On the stability of some controlled Markov chains and its applications to stochastic approximation with Markovian dynamic

Christophe Andrieu, Vladislav B. Tadić and Matti Vihola

School of Mathematics, University of Bristol

Department of Mathematics and Statistics, University of Jyväskylä

Abstract

We develop a practical approach to establish the stability, that is the recurrence in a given set, of a large class of controlled Markov chains. These processes arise in various areas of applied science and encompass in particular important numerical methods. We show in particular how individual Lyapunov functions and associated drift conditions for the parametrised family of Markov transition probabilities and the parameter update can be combined to form Lyapunov functions for the joint process, leading to the proof of the desired stability property. Of particular interest is the fact that the approach applies even in situations where the two components of the process present a time-scale separation, which is a crucial feature of practical situations. We then move on to show how such a recurrence property can be used in the context of stochastic approximation in order to prove the convergence of the parameter sequence, including in the situation where the so-called stepsize is adaptively tuned. We finally show that the results apply to various algorithms of interest in computational statistics and cognate areas.

1 Introduction: recurrence of controlled MC and compound drifts

The class of controlled Markov chain processes underpins numerous models or algorithms encountered in various areas of engineering or science (control, EM algorithm, adaptive MCMC). Consider the space $(\mathsf{X},\mathcal{B}(\mathsf{X}))$ where $\mathsf{X}\subset\mathbb{R}^{n_x}$ for some $n_x\geq 1$, a parametrized family of Markov transition probabilities $\{P_\theta,\theta\in\Theta\}$ (for some set $\Theta\subset\mathbb{R}^{n_\theta}$) such that for any $\theta,x\in\Theta\times\mathsf{X}$, $P_\theta(x,\cdot)$ is a probability distribution on $(\mathsf{X},\mathcal{B}(\mathsf{X}))$. The class of controlled Markov chains we consider in this paper consists of the class of processes defined on $\left((\Theta\times\mathsf{X})^{\mathbb{N}},(\mathcal{B}(\Theta)\otimes\mathcal{B}(\mathsf{X}))^{\otimes\mathbb{N}}\right)$ initialised at some $(\theta_0,X_0)=(\theta,x)\in\Theta\times\mathsf{X}$, with probability distribution denoted $\mathbb{P}_{\theta,x}(\cdot)$ (and associated expectation $\mathbb{E}_{\theta,x}(\cdot)$) and defined recursively for $i\geq 0$ as follows,

$$\begin{aligned} &X_{i+1}|(\theta_0,X_0,X_1,\ldots,X_i)\sim P_{\theta_i}(X_i,\cdot)\\ &\theta_{i+1}:=\phi_{i+1}(\theta_0,X_0,X_1,\ldots,X_{i+1}) \quad , \end{aligned} \tag{1.1}$$

for a family of mappings $\{\phi_i:\Theta\times\mathsf{X}^{i+1}\to\Theta\}$. The present paper is concerned with the stability of the sequence $\{\theta_i,X_i\}$, or more precisely the recurrence of such a process in a set $\mathcal{C}\subset\mathsf{X}\times\Theta$ i.e. we aim to develop practically relevant tools to establish that $\{\theta_i,X_i\}$ visits $\mathcal{C}$ infinitely often $\mathbb{P}_{\theta,x}-$a.s. Such a form of stability is central to establish important properties of the process which, depending on the context, range from the existence of an invariant distribution for the process or its marginals to the convergence of the parameter sequence $\{\theta_i\}$ to a set of values of particular interest. This is largely an open problem despite its practical relevance as illustrated and discussed later in the paper. The following toy example illustrates the potential difficulties one may face. Let $\mathsf{X}=\{0,1\}$ and consider the transition matrix

$$P_\theta=\left[\begin{array}{cc} 1-\exp(-|\theta|) & \exp(-|\theta|)\\ \exp(-|\theta|) & 1-\exp(-|\theta|)\end{array}\right],$$

with $\Theta=\mathbb{R}$. This transition matrix has $\pi=(1/2,1/2)$ as invariant distribution and its second eigenvalue is $\lambda=1-2\exp(-|\theta|)$. Set $\theta_{i+1}=\theta_i+a/i[1/2-X_{i+1}]$ for some $a>0$. One could expect $\{\theta_i\}$ to converge to

a finite value, but following the argument of [15, Section 6.3] one can in fact show that for some values of a, with positive probability $\{X_i\}$ may get stuck in either states while $\{\theta_i\}$ diverges. Ergodicity is lost here due to the fact that $\mathcal{C}_0 = \Theta \times \{0\}$ or $\mathcal{C}_1 = \Theta \times \{1\}$ is not visited infinitely often with probability one.

The remainder of the paper is organised as follows. In Section 2 we introduce our methodology, which relies on a classical Lyapunov function / drift condition to establish recurrence to a set. Of particular interest is that we show how individual drifts characterising the evolution from X_i to X_{i+1} and θ_i to θ_{i+1} can be combined to usefully characterise the joint dynamic, even in situation where this dynamic exhibits time-scale separation. In Section 3 we show how such a result can be established for a class of processes which covers numerous applications and in Section 4 we present generic applications of our results in the context of stochastic approximation [16]. In Section 5 we show how the results apply to both the AM algorithm of [10] but also the coerced acceptance probability algorithm [4, 6] and a novel variation.

2 Compound Lyapunov functions for some two timescale controlled Markov chains

The approach we adopt throughout this paper relies on a classical Lyapunov function and drift argument commonly used in the (homogeneous) Markov chain setting [12]. Due to the potential time inhomogeneity of the process above it is useful to consider a sequence of Lyapunov functions $\{W_i\}$ satisfying a sequence of drift conditions and leading to the following classical result, provided here together with its proof for completeness only. Hereafter, for any $i \geq 0$ we let $\mathcal{F}_i := \sigma(\theta_0, X_0, X_1, \ldots, X_i)$.

Lemma 1. *Let $\{W_i\}$ be a sequence of functions $W_i : \Theta \times \mathsf{X} \to [0, \infty)$ such that for the controlled Markov chain defined in (1.1) for all $\theta, x \in \Theta \times \mathsf{X}$*

1. for all $i \geq 0$, $\mathbb{E}_{\theta,x}[W_i(\theta_i, X_i)] < \infty$,
2. there exist $\mathcal{C} \subset \Theta \times \mathsf{X}$, a sequence $\{\delta_i, i \geq 1\}$ of non-negative scalars such that $\sum_{i=1}^{\infty} \delta_i = \infty$ and an integer $i_w < \infty$ such that for all $i \geq i_w$, and whenever $(\theta_i, X_i) \notin \mathcal{C}$, $\mathbb{P}_{\theta,x}$−a.s.

$$\mathbb{E}_{\theta,x}[W_{i+1}(\theta_{i+1}, X_{i+1}) \mid \mathcal{F}_i] \leq W_i(\theta_i, X_i) - \delta_{i+1} \quad . \tag{2.1}$$

Then $\sum_{i=1}^{n} \mathbb{I}\{(\theta_i, X_i) \in \mathcal{C}\} = \infty$, $\mathbb{P}_{\theta,x}$−a.s.

Proof. For any $k \geq 1$ we introduce the stopping times $\tau(k) := \inf\{i > k : (\theta_i, X_i) \in \mathcal{C}\}$. We proceed by contradiction and observe first that if the claim did not hold, then there would be an integer $i_w \leq n < \infty$ such that with positive probability the stopping time $\tau(n)$ would be infinite, i.e. $\mathbb{P}_{\theta,x}(\tau(n) = \infty) > 0$. We establish a result similar to [12, Proposition 11.3.3, p. 266], but take care of the inhomogeneity and do not require the same precision. We introduce the following notation for simplicity: $W_i := W_i(\theta_i, X_i)$ and for any $m \in \mathbb{N}$, $\tau^m := \tau(n) \wedge m$ (we omit the dependence on n in order to alleviate notation). Assumption (2.4) implies that for $i \geq n+1$,

$$\begin{aligned}\mathbb{E}_{\theta,x}[W_{i+1}\mathbb{I}\{\tau^m \geq i+1\}] &= \mathbb{E}_{\theta,x}[W_i\mathbb{I}\{\tau^m \geq i+1\} + \mathbb{E}_{\theta,x}\left[W_{i+1} - W_i \mid \mathcal{F}_i]\mathbb{I}\{\tau^m \geq i+1\}\right] \\ &\leq \mathbb{E}_{\theta,x}[W_i\mathbb{I}\{\tau^m \geq i\}] - \mathbb{E}_{\theta,x}[\delta_{i+1}\mathbb{I}\{\tau^m \geq i+1\}] \quad .\end{aligned}$$

and consequently we can establish

$$\mathbb{E}_{\theta,x}[\textstyle\sum_{i=n+1}^{\infty}\delta_{i+1}\mathbb{I}\{\tau^m - 1 \geq i\}] \leq \mathbb{E}_{\theta,x}[W_{n+1}] - \mathbb{E}_{\theta,x}[W_{\tau^m}] \leq \mathbb{E}_{\theta,x}[W_{n+1}] \quad .$$

Now, by using the trivial inequality $\mathbb{E}_{\theta,x}[\mathbb{I}\{\tau(n) = \infty\}\sum_{i=n+1}^{\infty}\delta_{i+1}\mathbb{I}\{\tau^m - 1 \geq i\}] \leq \mathbb{E}_{\theta,x}[\sum_{i=n+1}^{\infty}\delta_{i+1}\mathbb{I}\{\tau^m - 1 \geq i\}]$ and the monotone convergence theorem (thanks to our assumptions on $\{\delta_i\}$) we obtain the contradictory statement

$$\mathbb{P}_{\theta,x}(\tau(n) = \infty) \sum_{i=n+1}^{\infty} \delta_{i+1} \leq \mathbb{E}_{\theta,x}[W_{n+1}] < \infty \quad .$$

We therefore conclude that for any $i \geq i_w$, $\mathbb{P}_{\theta,x}(\tau(i) = \infty) = 0$ and the result follows. □

The main result of this section consists of showing that it is possible to construct joint Lyapunov function sequences $\{W_i\}$ which satisfy drifts to a set $\mathcal{C}$, such that the conditions of Lemma 1 hold, from two separate Lyapunov functions $w(\theta)$ and $V(x)$ each satisfying an individual drift condition characterising the two respective updates involved in the definition of $\{\theta_i, X_i\}$ in (1.1). The form of these individual drifts is given below in (2.2) and (2.3) : it is worth pointing out that we allow the drift on $w(\theta)$ to vanish with time since $\{\gamma_i\}$ may be allowed to vanish. This is practically very relevant since in many situations of interest the "size" of the increments $|\theta_{i+1} - \theta_i|$ may vanish as $i \to \infty$ while that of $|X_{i+1} - X_i|$ may not. The role of the sequence $\{\gamma_i\}$ is to accommodate the possibility of two distinct timescales for the two updates in (1.1) - examples are numerous and some will be presented later in Sections 4 and 5. We will consider two scenarios which share very similar assumptions, and will be labelled with $s \in \{0, 1\}$.

(**A1**) Suppose $V : \mathsf{X} \to [1, \infty)$ and $w : \Theta \to [1, \infty)$ are two functions such that there exist functions $\Delta_w, \Delta_V : \Theta \times \mathsf{X} \to \mathbb{R}$, a set $\mathcal{C} \subset \Theta \times \mathsf{X}$, a sequence of strictly positive integers $\{\gamma_i, i \geq 1\}$ such that

1. $\{\gamma_i\}$ is bounded,
2. for some integer $i_0 \geq 0$, $\mathbb{P}_{\theta,x}-$a.s. the following individual drifts hold for all $i \geq i_0$,

$$\mathbb{E}_{\theta,x}[w(\theta_{i+1}) \mid \mathcal{F}_i] \leq w(\theta_i) - \gamma_{i+1}\Delta_w(\theta_i, X_i) \tag{2.2}$$

$$\mathbb{E}_{\theta,x}[V(X_{i+1}) \mid \mathcal{F}_i] \leq V(X_i) - \Delta_V(\theta_i, X_i) \quad , \tag{2.3}$$

   and $\mathbb{E}_{\theta,x}[w(\theta_i)] < \infty$ and $\mathbb{E}_{\theta,x}[V(X_i)] < \infty$,
3. there exist constants $\delta \in (0, \infty)$ and $\upsilon_v, \upsilon_w \in (0, 1]$ such that

$$\upsilon_w \frac{\Delta_w(\theta, x)}{w^{1-\upsilon_w}(\theta)} + \upsilon_v \frac{\Delta_V(\theta, x)}{V^{1-\upsilon_v}(x)} \geq \delta w^{s \times \upsilon_w}(\theta) \quad \text{for} \quad (\theta, x) \notin \mathcal{C} \tag{2.4}$$

   and

$$\sup_{(\theta,x)\in\mathcal{C}} |\Delta_w(\theta, x)| \vee |\Delta_V(\theta, x)| < \infty \quad .$$

The following theorem establishes two recurrence results for $\{\theta_i, X_i\}$ to $\mathcal{C}$. The first result requires the strongest set of assumptions but also establishes a stronger result, namely that the first moment of the return times to $\mathcal{C}$ are uniformly bounded in time. The second result requires weaker assumptions but does not guarantee the existence of a uniform in time upper bound on characteristics of the return times. A particular contribution here is the rescaling of either the Lyapunov function $w(\theta)$ or $V(x)$ in order to allow for their respective drift terms to be compared on the same time scale.

Theorem 1. *Consider the controlled Markov chain defined in (1.1). Define the sequences of functions $\{W_i : \Theta \times \mathsf{X} \to [1, \infty)\}$ and $\{U_i : \Theta \times \mathsf{X} \to [1, \infty\}$ for $i \geq 1$ and $\theta, x \in \Theta \times \mathsf{X}$ as follows*

$$W_i(\theta, x) := V^{\upsilon_v}(x) + w^{\upsilon_w}(\theta)/\gamma_i \quad \text{and} \quad U_i(\theta, x) := \gamma_i W_i(\theta, x) \quad ,$$

where $\{\gamma_i\}$, $w(\cdot)$, $V(\cdot)$, υ_v and υ_w are as in (A1), which is assumed to hold. Then,

1. *if $s = 1$ and $\bar{\ell} := \limsup_{i\to\infty}(\gamma_{i+1}^{-1} - \gamma_i^{-1}) < \delta$, then for any $\delta_W \in (0, \delta - \bar{\ell})$ there exists $i_W \geq i_0$ such that for any $i \geq i_W$, whenever $(\theta_i, X_i) \notin \mathcal{C}$, $\mathbb{P}_{\theta,x}-$a.s.*

$$\mathbb{E}_{\theta,x}[W_{i+1}(\theta_{i+1}, X_{i+1}) \mid \mathcal{F}_i] \leq W_i(\theta_i, X_i) - \delta_W \quad , \tag{2.5}$$

   and $\mathbb{E}_{\theta,x}[W_i(\theta_i, X_i)] < \infty$, and $\sum_{i=1}^{\infty} \mathbb{I}\{(\theta_i, X_i) \in \mathcal{C}\} = \infty$, $\mathbb{P}_{\theta,x}-$almost surely

2. *if $s = 0$, $\{\gamma_i\}$ is non-increasing then for any $i \geq i_0$, whenever $(\theta_i, X_i) \notin \mathcal{C}$, $\mathbb{P}_{\theta,x}-$a.s.*

$$\mathbb{E}_{\theta,x}[U_{i+1}(\theta_{i+1}, X_{i+1}) \mid \mathcal{F}_i] \leq U_i(\theta_i, X_i) - \delta\gamma_{i+1} \tag{2.6}$$

   and moreover $\mathbb{E}_{\theta,x}[U_i(\theta_i, X_i)] < \infty$. If in addition $\sum_{i=1}^{\infty} \gamma_i = \infty$, then $\sum_{i=1}^{\infty} \mathbb{I}\{(\theta_i, X_i) \in \mathcal{C}\} = \infty$, $\mathbb{P}_{\theta,x}-$almost surely.

Proof. We start with the scenario where $s = 1$. By (A1), Jensen's inequality and the classical concavity identity $(1+x)^{\upsilon} \leq 1 + \upsilon x$ for $x \in [-1, \infty)$ and $\upsilon \in (0,1]$, we have for any $i \geq i_0$ and $\mathbb{P}_{\theta,x}-$a.s.

$$\begin{aligned}\mathbb{E}_{\theta,x}[W_{i+1}(\theta_{i+1}, X_{i+1}) \mid \mathcal{F}_i] \leq & V^{\upsilon_v}(X_i)\left(1 - \frac{\Delta_V(\theta_i, X_i)}{V(X_i)}\right)^{\upsilon_v} + \gamma_{i+1}^{-1} w^{\upsilon_w}(\theta_i)\left(1 - \gamma_{i+1}\frac{\Delta_w(\theta_i, X_i)}{w(\theta_i)}\right)^{\upsilon_w} \\ \leq & V^{\upsilon_v}(X_i)\left(1 - \upsilon_v \frac{\Delta_V(\theta_i, X_i)}{V(X_i)}\right) + \gamma_{i+1}^{-1} w^{\upsilon_w}(\theta_i)\left(1 - \gamma_{i+1}\upsilon_w \frac{\Delta_w(\theta_i, X_i)}{w(\theta_i)}\right) \\ = & W_i(\theta_i, X_i) + \left(\gamma_{i+1}^{-1} - \gamma_i^{-1}\right) w^{\upsilon_w}(\theta_i) - \left(\upsilon_v \frac{\Delta_V(\theta_i, X_i)}{V^{1-\upsilon_v}(X_i)} + \upsilon_w \frac{\Delta_w(\theta_i, X_i)}{w^{1-\upsilon_w}(\theta_i)}\right)\end{aligned}$$

Let $\delta_W \in (0, \delta - \bar{\ell})$ and $i_W \geq i_0$ be such that $\sup_{i \geq i_w}(\gamma_{i+1}^{-1} - \gamma_i^{-1}) < \delta - \delta_W$. Then, for all $i \geq i_W$ and $(\theta_i, X_i) \notin \mathcal{C}$, $\mathbb{P}_{\theta,x}-$a.s.

$$\mathbb{E}_{\theta,x}[W_{i+1}(\theta_{i+1}, X_{i+1}) \mid \mathcal{F}_i] \leq W_i(\theta_i, X_i) - \delta_W \quad .$$

Let $C := [\sup_{i \geq i_0} \gamma_i(\gamma_{i+1}^{-1} - \gamma_i^{-1})] \vee [\sup_{(\theta,x)\in\mathcal{C}} |\gamma_1 \Delta_w(\theta, x)| \vee |\Delta_V(\theta, x)|]$. Now for any $i \geq i_0$ and $(\theta_i, X_i) \in \Theta \times \mathsf{X}$ we have, starting as above,

$$\begin{aligned}\mathbb{E}_{\theta,x}[W_{i+1}(\theta_{i+1}, X_{i+1}) \mid \mathcal{F}_i] &\leq (1+C)^{\upsilon_v} V^{\upsilon_v}(X_i) + (1+C)^{\upsilon_w}[\gamma_i(\gamma_{i+1}^{-1} - \gamma_i^{-1}) + 1] w^{\upsilon_w}(\theta_i)/\gamma_i \\ &\leq (1+C)^2 W_i(\theta_i, X_i) \quad .\end{aligned}$$

From these inequalities we therefore deduce that for any $i \geq i_0$, $\mathbb{E}_{\theta,x}[W_i(\theta_i, X_i)] \leq (1+C)^{2(i-i_0)}\mathbb{E}_{\theta,x}[W_{i_0}(\theta_{i_0}, X_{i_0})] < \infty$ where the last inequality follows from our assumptions. For the scenario where $s = 0$ with $U_i(\theta, x) = \gamma_i W_i(\theta, x)$ we obtain from above

$$\begin{aligned}\mathbb{E}_{\theta,x}[U_{i+1}(\theta_{i+1}, X_{i+1}) \mid \mathcal{F}_i] \leq & U_i(\theta_i, X_i) + (\gamma_{i+1} - \gamma_i) W_i(\theta_i, X_i) \\ & + \gamma_{i+1}\left(\gamma_{i+1}^{-1} - \gamma_i^{-1}\right) w^{\upsilon_w}(\theta_i) - \gamma_{i+1}\left(\upsilon_v \frac{\Delta_V(\theta_i, X_i)}{V^{1-\upsilon_v}(X_i)} + \upsilon_w \frac{\Delta_w(\theta_i, X_i)}{w^{1-\upsilon_w}(\theta_i)}\right)\end{aligned}$$

and since

$$(\gamma_{i+1} - \gamma_i) W_i(\theta, x) + \gamma_{i+1}\left(\gamma_{i+1}^{-1} - \gamma_i^{-1}\right) w^{\upsilon_w}(\theta) = (\gamma_{i+1} - \gamma_i) V(x)$$

which together with the fact that $\{\gamma_i\}$ is non-increasing leads to (2.6) for $\theta_i, X_i \in \mathcal{C}^c$. Notice further that $U_i(\theta, x) \leq \gamma_1 W_i(\theta, x)$, implying $\mathbb{E}_{\theta,x}[U_i(\theta_i, X_i)] < \infty$ for any $i \geq i_0$. We now conclude in both scenarios with Lemma 1. □

Some comments are in order concerning the choice of the Lyapunov functions and the assumptions. First we clarify the role of υ_v and υ_w, which are additional degrees of freedom one may find helpful to establish (2.4) in regions of $\Theta \times \mathsf{X}$ where $\Delta_V(\theta, x)$ (resp. $\Delta_w(\theta, x)$) is negative and of large magnitude but V (resp. w) is itself large. Notice also that more general concave transformations of V and w could be considered for the definition of W_i and U_i, but we do not pursue this here. We would also like to point out that other Lyapunov functions of the form $U_i^{\alpha}(\theta, x) := \gamma_i^{\alpha} W_i(\theta, x)$ for $\alpha \geq 0$ may be considered but we have found the scenarios $\alpha = 0$ and $\alpha = 1$ to be of interest only. Finally whereas it is clear that (2.4) is stronger for $s = 1$ than $s = 0$, we also note that $\limsup_{i\to\infty}(\gamma_{i+1}^{-1} - \gamma_i^{-1}) < \delta - \delta_W$ implies $\sum_{i=1}^{\infty} \gamma_i = \infty$.

In the next section we consider a practically relevant scenario encountered in practice, for which we identify $\mathcal{C}$ and $\{W_i\}$, but also establish an even stronger drift than in (2.5). We will show in Section 5 that such results are satisfied in realistic scenarios.

3 Simultaneous $\theta-$dependent drift conditions and stability

The results presented in the previous section are rather abstract since neither of Δ_w, Δ_V and $\mathcal{C}$ are specified. Here we add some structure, and in particular show how a simultaneous drift condition on the family of Markov transition probabilities P_θ for $\theta \in \Theta$, where the dependence on θ is explicit, can be used to prove the stability of the sequence $\{\theta_i, X_i\}$ to a well identified set $\mathcal{C} \subset \Theta \times \mathsf{X}$. For ease of exposition we focus throughout this section on the situation where $\phi_i := \phi_{\gamma_i}$ for some family of updates $\{\phi_\gamma : \Theta \times \mathsf{X} \to \Theta, \gamma \in (0, \gamma^+]\}$ and a positive sequence $\{\gamma_i\} \subset (0, \gamma^+]^{\mathbb{N}}$, allowing us to define the update $\theta_{i+1} = \phi_{\gamma_{i+1}}(\theta_i, X_{i+1})$ for $i \geq 0$. This directly covers most relevant applications in computational statistics and can be easily generalized. As we shall see the realistic assumptions we use lead in fact to stronger results than those of the previous section. For any $f : \mathsf{X} \to \mathbb{R}$ we use the standard notation $P_\theta f(x) := \int_{\mathsf{X}} P_\theta(x, dy) f(y)$. The $\theta-$dependent simultaneous drift conditions we consider here are as follows.

(A2) The family or Markov transition probabilities $\{P_\theta, \theta \in \Theta\}$ is such that there exist

1. $V : \mathsf{X} \to [1, +\infty)$ and $\mathsf{C} \subset \mathsf{X}$ such that $\sup_{x \in \mathsf{C}} V(x) < +\infty$,
2. $a(\cdot), b(\cdot) : \Theta \to [0, +\infty)$ and $\iota \in [0, 1]$

such that for any $\theta, x \in \Theta \times \mathsf{X}$

$$P_\theta V(x) \le \left[V(x) - a^{-1}(\theta) V^\iota(x)\right] \mathbb{I}\{x \notin \mathsf{C}\} + b(\theta)\mathbb{I}\{x \in \mathsf{C}\} \ .$$

We define the level sets of $a(\cdot)$ (resp. $w(\cdot)$): for all $M \ge 0$, $\mathcal{A}_M := \{\theta \in \Theta : a(\theta) \le M\}$ (resp. $\mathcal{W}_M := \{\theta \in \Theta : w(\theta) \le M\}$) and for any set A we will denote A^c the complement of A in either Θ or X. Notice that assumption (A2) implies that $\inf_{\theta \in \Theta} a(\theta) > 0$. The situations we are interested in are those for which $\mathcal{A}_M^c \ne \emptyset$ for any $M > 0$. Hereafter it will be convenient to denote for any $\theta, x \in \Theta \times \mathsf{X}$, $\gamma \in (0, \gamma^+]$ and $f : \Theta \times \mathsf{X} \to \mathbb{R}^{n_f}$

$$P_{\theta,\gamma} f(\theta, x) := \int_{\mathsf{X}} P_\theta(x, dy) f(\phi_\gamma(\theta, y), y) \quad ,$$

and for any $u, v \in \mathbb{R}^2$ we define $u \vee v := \max\{u, v\}$ and $u \wedge v := \min\{u, v\}$.

(A3) The family of mappings $\{\phi_\gamma : \Theta \times \mathsf{X} \to \Theta, \gamma \in (0, \gamma^+]\}$ is such that there exists a Lyapunov function $w : \Theta \to [1, \infty)$ such that (with $\{P_\theta, \theta \in \Theta\}$, $V(\cdot), \mathsf{C}, a(\cdot), b(\cdot)$ and ι as in (A2))

1. $\lim_{M \to \infty} \inf_{\theta \in \mathcal{A}_M^c} w(\theta) = \infty$ and $\lim_{M \to \infty} \sup_{\theta \in \mathcal{W}_M^c} b(\theta)/w(\theta) = 0$
2. there exists $\beta \in [0, 1]$ and $c(\cdot), d(\cdot) : \Theta \to [0, \infty)$ such that for all $\gamma \in (0, \gamma^+]$ and $\theta, x \in \Theta \times \mathsf{X}$,

   $$P_{\theta,\gamma} w(\theta, x) \le w(\theta) - \gamma w(\theta) \Delta \left(V_\beta(\theta, x) \mathbb{I}\{x \notin \mathsf{C}\} + d(\theta)\mathbb{I}\{x \in \mathsf{C}\}\right) \quad ,$$

   where $V_\beta(\theta, x) := c(\theta) + V^\beta(x)/w(\theta)$ and where
3. $c(\cdot), d(\cdot) : \Theta \to [0, \infty)$ are bounded and $\lim_{M \to \infty} \sup_{\theta \in \mathcal{W}_M^c} [c(\theta) \vee d(\theta)] = 0$,
4. $\Delta(\cdot) : [0, \infty) \to \mathbb{R}$ is such that
   (a) $\Delta(0) > 0$ and it is continuous in a neighbourhood of 0,
   (b) there exists $p_\Delta \in (0, \iota/\beta]$ such that for all $M > 0$ there exists $C_{\Delta,M} > 0$ such that for all $z \ge M$

   $$|\Delta(z)| \le C_{\Delta,M} \times z^{p_\Delta} \quad ,$$

5. for any $\epsilon > 0$

   $$\sup_{\theta, x \in \tilde{\mathcal{V}}_\epsilon} \frac{a(\theta) w^{1-p_\Delta}(\theta)}{V^{\iota - p_\Delta \beta}(x)} < \infty \quad ,$$

   where $\tilde{\mathcal{V}}_\epsilon := \{\theta, x : V^\beta(x)/w(\theta) \ge \epsilon\}$.

Remark 1. The conditions above may appear abstract, but are motivated by the following concrete situations:

1. The simultaneous fixed$-\theta$ drift conditions (A2) can be established in numerous situations of practical interest. Examples are given in Section 5, where the transition probabilities share the same invariant distribution, but it should be pointed out that such drift conditions can also be established in situations of interest where each transition kernel P_θ has its own invariant distribution π_θ; this is the case for example in the context of the stochastic approximation implementation of the EM algorithm in [5]. Other examples can be found in[16] for algorithms used in the area of digital communications, although the dependence on θ is never used.

2. Typically the function $\Delta(\cdot)$ in (A3) will take the form of a polynomial, as a byproduct of a tractable approximation of $w(\phi_\gamma(\theta, y))$ in terms of $w(\theta)$. For example in the situation where $\vartheta = \phi_\gamma(\theta, y) = \theta + \gamma H(\theta, y)$, which corresponds to the standard stochastic approximation framework (see Section 4), a Taylor expansion of $w(\vartheta)$ around θ will lead to,

   $$w(\vartheta) \le w(\theta) + \gamma \langle H(\theta, y), \nabla w(\theta)\rangle + \frac{1}{2}\gamma^2 \overline{w}'' \times |H(\theta, y)|^2$$

   whenever $\bar{w}'' := \sup_{\theta \in \Theta} |\nabla^2 w(\theta)| < \infty$. With appropriate assumptions on $H(\theta, y)$ one can apply P_θ to both sides of this inequality and hence obtain a drift condition on $w(\cdot)$ of the form given in (A3).

3. The condition required on $p_\Delta \in [0, \iota/\beta]$ can be understood as being a tradeoff between the strength of the drift in (A2) and the strength of unfavorable updates $\theta_+ = \phi_\gamma(\theta, x_+)$ such that $w(\theta_+) \gg w(\theta)$.

The following proposition allows one to check (A3)-5 easily in numerous situations.

Proposition 1. *Let $a(\cdot), w(\cdot), \beta$ and ι be as in (A3). Assume that there exists $C > 0$ and $\varkappa \geq 0$ such that for any $\theta \in \Theta$, $a(\theta) \leq C w^{\varkappa}(\theta)$ and $\beta \in [0, \iota/(1+\varkappa)]$. Then (A3)-5 holds for $V(\cdot)$ as in (A3).*

Proof. Consider first the case $p_\Delta \leq 1 + \varkappa$ and notice that $\iota - \beta p_\Delta \geq (1 + \varkappa - p_\Delta)\beta$. This leads for $(\theta, x) \in \tilde{\mathcal{V}}_\epsilon$ to

$$\frac{a(\theta) w^{1-p_\Delta}(\theta)}{V^{\iota - \beta p_\Delta}(x)} \leq C \frac{w^{1+\varkappa-p_\Delta}(\theta)}{V^{\iota-\beta p_\Delta}(x)} \leq C \left(\frac{w(\theta)}{V^{\beta}(x)}\right)^{1+\varkappa-p_\Delta} \leq C \epsilon^{-(1+\varkappa-p_\Delta)} \quad .$$

Now assume that $p_\Delta > 1 + \varkappa$. Since $w \geq 1$, $V \geq 1$ and $\iota \geq \beta p_\Delta$, using the first inequality above we deduce that

$$\frac{a(\theta) w^{1-p_\Delta}(\theta)}{V^{\iota-\beta p_\Delta}(x)} \leq C \quad .$$

□

Hereafter for any $\varepsilon \in (0, \Delta(0))$ we will denote

$$\Gamma_\varepsilon := \left\{\gamma, \bar{\gamma} \in (0, \gamma^+] : 0 \leq \gamma^{-1} - \bar{\gamma}^{-1} < \Delta(0) - \varepsilon\right\} \ ,$$

where we omit the dependence on γ^+ for simplicity.

Theorem 2. *Assume that $\{P_\theta, \theta \in \Theta\}$ and $\{\phi_\gamma, \gamma \in (0, \gamma^+]\}$ satisfy (A2) and (A3). Then for any $\varepsilon \in (0, \Delta(0))$ there exist $\lambda_* \in [1, \infty)$, $\delta, M_* \in (0, +\infty)$ such that for any $\gamma, \bar{\gamma} \in \Gamma_\varepsilon$ and $\theta, x \notin \mathcal{W}_{M_*} \times \mathsf{C}$,*

$$P_{\theta,\gamma}\{\lambda_* V + w/\gamma\}(\theta, x) \leq \lambda_* V(x) + w(\theta)/\bar{\gamma} - \delta[V^{\iota}(x)/a(\theta) + w(\theta)] \tag{3.1}$$

Corollary 1. *Let $\{\theta_i, X_i\}$ be the controlled Markov chain process as described in Eq. (1.1) with for any $i \geq 1$ $\phi_i(\theta_0, x_0, x_1, \ldots, x_i) := \phi_{\gamma_i}(\theta_{i-1}, x_i)$ for a family $\{\phi_\gamma : \Theta \times \mathsf{X} \to \Theta, \gamma \in (0, \gamma^+]\}$ and some real positive sequence $\{\gamma_i\}$. Assume further that $\{P_\theta, \theta \in \Theta\}$ and $\{\phi_\gamma, \gamma \in (0, \gamma^+]\}$ satisfy (A2), (A3) and that $\{\gamma_i\}$ is such that*

$$\limsup_{i\to\infty} (\gamma_{i+1}^{-1} - \gamma_i^{-1}) < \Delta(0) \ . \tag{3.2}$$

Then, with M_* as in Theorem 2 the set $\mathcal{W}_{M_*} \times \mathsf{C}$ is visited infinitely often $\mathbb{P}_{\theta,x}-$a.s. by $\{\theta_i, X_i\}$.

Proof. (Corollary 1). Let $\delta \in (0, 1]$, $\lambda_* > 1$ and $M_* > 0$ be as in Theorem 2 and define the family of (Lyapunov) functions $\{W_i(\theta, x) := \lambda_* V(x) + w(\theta)/\gamma_i\}$. From the assumption on $\{\gamma_i\}$ there exists $i_0 \in \mathbb{N}$ such that for any $i \geq i_0$ and $\theta, x \notin \mathcal{W}_{M_*} \times \mathsf{C}$

$$P_{\theta,\gamma_i} W_i(\theta, x) \leq W_{i-1}(\theta, x) - \delta[V^{\iota}(x)/a(\theta) + w(\theta)] \ .$$

The result follows from Lemma 1 since $\inf_{\theta\in\Theta} w(\theta) > 0$. □

Remark 2. One can notice that,

1. in the case where $\gamma_i = c_0/(c_1 + i)^a$ (3.2) is satisfied for any $c_0 > 0$ and $a \in (0, 1)$, and for $c_0 < \Delta(0)$ when $a = 1$,

2. in the case where $\{\gamma_i = \gamma \leq \gamma^+\}$ is constant, $W_0(\theta, x) = \lambda_* V(x) + w(\theta)$ and for any $\theta \in \Theta$, $a(\theta) \leq C w^{\varkappa}(\theta)$ for $\varkappa > 0$ then one may show that for any $i \geq i_0$ and $\theta, x \notin \mathcal{W}_{M_*} \times \mathsf{C}$

$$P_{\theta,\gamma} W_0(\theta, x) \leq W_0(\theta, x) - \delta' W_0^{\iota/(1+\varkappa)}(\theta, x) \ .$$

Indeed, from a standard convexity inequality, for any $l \in (0, 1]$,

$$l \frac{V^{\iota}(x)}{w(\theta)^{\varkappa}} + (1 - l) w(\theta) \geq \left(\frac{V^{\iota}(x)}{w(\theta)^{\varkappa}}\right)^{l} w^{1-l}(\theta) \ . \tag{3.3}$$

which, with the choice $\bar{l} = 1/(1+\varkappa)$, leads to

$$V^{\iota}(x)/w^{\varkappa}(\theta) + w(\theta) \geq \bar{l} V^{\iota}(x)/w^{\varkappa}(\theta) + (1 - \bar{l}) w(\theta) \geq V^{\iota/(1+\varkappa)}(x) \quad .$$

As a result

$$\begin{aligned}V^\iota(x)/a(\theta)+w(\theta) &\geq \frac{1}{2}[V^{\iota/(1+\varkappa)}(x)+w(\theta)]\\ &\geq \frac{1}{2}[V^{\iota/(1+\varkappa)}(x)+w^{\iota/(1+\varkappa)}(\theta)]\\ &\geq 2^{-1-\iota/(1+\varkappa)}\left(V(x)+w(\theta)\right)^{\iota/(1+\varkappa)}\end{aligned}$$

and we conclude. This suggests the possibility to precisely characterise the return times to $\mathcal{W}_{M_*}\times\mathsf{C}$ as this form of drift condition is known to lead to the existence of polynomial moments of return times.

Proof. (Theorem 2) Choose $\varepsilon\in(0,\Delta(0))$ and $\epsilon_->0$ such that for any $|z|\leq\epsilon_-$, $|\Delta(0)-\Delta(z)|\leq\varepsilon/2$. This implies,

$$\sup_{\gamma,\bar\gamma\in\Gamma_\varepsilon}(\gamma^{-1}-\bar\gamma^{-1})-\inf_{\{z:|z|\leq\epsilon_-\}}\Delta(z)\leq\Delta(0)-\varepsilon+\varepsilon/2-\Delta(0)=-\varepsilon/2 \quad . \tag{3.4}$$

Now let $M_0\geq 0$ be such that $\sup_{\theta\in\mathcal{W}^c_{M_0}}d(\theta)\leq\epsilon_-$ and $\sup_{\theta\in\mathcal{W}^c_{M_0}}c(\theta)\leq\epsilon_-/2$. From (A2) and (A3) we have for $(\theta,x)\in\Theta\times\mathsf{X}$ and $\lambda\in(0,\infty)$

$$\begin{aligned}P_{\theta,\gamma}\{\lambda V+w/\gamma\}(\theta,x)\leq\lambda\left[V(x)-a^{-1}(\theta)V^\iota(x)\right]\mathbb{I}\{x\notin\mathsf{C}\}+\lambda b(\theta)\mathbb{I}\{x\in\mathsf{C}\}\\ +w(\theta)/\gamma-w(\theta)\Delta\left(V_\beta(\theta,x)\mathbb{I}\{x\notin\mathsf{C}\}+d(\theta)\mathbb{I}\{x\in\mathsf{C}\}\right)\end{aligned}$$

Note that for $(\theta,x)\in\mathcal{W}_{M_0}\times\mathsf{C}^c$, $V_\beta(\theta,x)\geq M_0^{-1}$ and therefore from (A3)-4b

$$w(\theta)\Delta(V_\beta(\theta,x))\leq C_{\Delta,M_0^{-1}}M_0\sup_{\theta\in\Theta}(w^{-1}(\theta)+c(\theta))^{p_\Delta}\times V^{p_\Delta\beta}(x) \quad ,$$

and $\sup_{\theta\in\mathcal{W}_{M_0}}\Delta(d(\theta))<\infty$ as $\Delta(\cdot)$ is bounded on compact sets and $\sup_{\theta\in\Theta}d(\theta)<\infty$. Let now

$$C'_{\Delta,M_0}:=[C_{\Delta,M_0^{-1}}M_0\sup_{\theta\in\Theta}(w^{-1}(\theta)+c(\theta))^{p_\Delta}]\vee[\sup_{\theta\in\mathcal{W}_{M_0}}\Delta(d(\theta))]<\infty \quad ,$$

notice that $\iota\geq p_\Delta\beta$ and recall that $V\geq 1$, then we have for $\gamma,\bar\gamma\in\Gamma_\varepsilon$

$$P_{\theta,\gamma}\{\lambda V+w/\gamma\}(\theta,x)\leq\lambda V(x)+w(\theta)/\bar\gamma+\Lambda(\theta,x) \quad ,$$

with

$$\begin{aligned}\Lambda(\theta,x):=-\lambda V(x)+\lambda\left[V(x)-a^{-1}(\theta)V^\iota(x)\right]\mathbb{I}\{x\notin\mathsf{C}\}+\lambda b(\theta)\mathbb{I}\{x\in\mathsf{C}\}\\ +(\Delta(0)-\varepsilon)w(\theta)-\Delta\left(V_\beta(\theta,x)\mathbb{I}\{x\notin\mathsf{C}\}+d(\theta)\mathbb{I}\{x\in\mathsf{C}\}\right)w(\theta)\ \mathbb{I}\{\theta\in\mathcal{W}^c_{M_0}\}\\ +C'_{\Delta,M_0}V^\iota(x)\ \mathbb{I}\{\theta\in\mathcal{W}_{M_0}\}\ .\end{aligned} \tag{3.5}$$

It will be convenient below to refer to the following inequality

$$\Lambda(\theta,x)\leq-\tilde\delta[V^\iota(x)/a(\theta)+w(\theta)]\ , \tag{3.6}$$

for $\theta,x\notin\mathcal{W}_{\tilde M}\times\mathsf{S}$ and various instantiations of $\tilde\delta,\tilde M,\lambda>0$ and $\mathsf{S}\subset\mathsf{X}$. Our ultimate aim is to prove that under the stated assumptions there exist $\delta,\lambda_*\in(0,+\infty)$ and $M_*\geq M_0$ such that (3.6) holds for $(\theta,x)\in(\mathcal{W}_{M_*}\times\mathsf{C})^c$. For any $M\geq M_0$ we use the following partition

$$(\mathcal{W}_M\times\mathsf{C})^c=(\mathcal{W}_{M_0}\times\mathsf{C}^c)\cup(\mathcal{W}^c_{M_0}\times\mathsf{C}^c)\cup(\mathcal{W}^c_M\times\mathsf{C})\ ,$$

which leads us to consider three cases, (a), (b) and (c) from left to right.

(a) For $(\theta,x)\in\mathcal{W}_{M_0}\times\mathsf{C}^c$ and any $\lambda>0$ we have

$$\begin{aligned}\Lambda(\theta,x)&\leq[\Delta(0)-\varepsilon]w(\theta)-\lambda V^\iota(x)/a(\theta)+C'_{\Delta,M_0}V^\iota(x)\ .\\ &\leq[\Delta(0)-\varepsilon]\sup_{\theta\in\mathcal{W}_{M_0}}w(\theta)+V^\iota(x)\left[C'_{\Delta,M_0}-\lambda/\sup_{\vartheta\in\mathcal{W}_{M_0}}a(\vartheta)\right]\ ,\end{aligned}$$

where we note that $\sup_{\vartheta\in\mathcal{W}_{M_0}} a(\vartheta) < \infty$ from (A3)-1. Now, from our choice of M_0 and since $V \geq 1$ and $\inf_{\vartheta\in\Theta} a(\vartheta) > 0$, we conclude about the existence of $\lambda_a, \delta_a > 0$ such that for all $\lambda \geq \lambda_a$, $(\theta, x) \in \mathcal{W}_{M_0} \times \mathsf{C}^c$

$$\Lambda(\theta, x) \leq [\Delta(0) - \varepsilon]M_0 + V^\iota(x) \left[C'_{\Delta,M_0} - \lambda / \sup_{\vartheta\in\mathcal{W}_{M_0}} a(\vartheta)\right] \leq -\delta_a[V^\iota(x)/a(\theta) + w(\theta)] \ .$$

Therefore (3.6) is satisfied with $\tilde{M} = M_0$, any $\lambda \geq \lambda_a$ and $\tilde{\delta} = \delta_a$.

(**b**) For $(\theta, x) \in \mathcal{W}^c_{M_0} \times \mathsf{C}^c$ and any $\lambda > 0$ we have

$$\Lambda(\theta, x) \leq -\lambda V^\iota(x)/a(\theta) + [\Delta(0) - \varepsilon - \Delta(V_\beta(\theta, x))]\, w(\theta) \ ,$$

and we seek to show that there exists $\lambda_b = \lambda_i \vee \lambda_{ii}$ and $\delta_b = \delta_i \wedge \delta_{ii} > 0$ (where $\lambda_i, \lambda_{ii} > 0$ and $\delta_i, \delta_{ii} > 0$ are given below in the proof) such that for all $\lambda \geq \lambda_b$ and $(\theta, x) \in \mathcal{W}^c_{M_0} \times \mathsf{C}^c$ (3.6) is satisfied with $\tilde{\delta} = \delta_b$. In what follows we will use the following intermediate results. From (A3)-5 we have that for any $(\theta, x) \in \mathcal{W}^c_{M_0} \times \mathsf{C}^c$ the condition

$$V_\beta(\theta, x) = \frac{V^\beta(x)}{w(\theta)} + c(\theta) \geq \epsilon_-, \text{ implies } \frac{V^\beta(x)}{w(\theta)} \geq \epsilon_- - \sup_{\vartheta\in\mathcal{W}_{M_0^c}} c(\vartheta) \geq \epsilon_-/2,$$

and therefore that for $q \in \{0, p_\Delta\}$

$$\sup_{(\mathcal{W}^c_{M_0}\times\mathsf{C}^c)\cap\{\theta,x: V_\beta(\theta,x)\geq\epsilon_-\}} \frac{a(\theta)w^{1-q}(\theta)}{V^{\iota-q\beta}(x)} \leq C_{\epsilon_-} < \infty \quad . \tag{3.7}$$

Indeed the case $q = p_\Delta$ is true by assumption and for $V^\beta(x)/w(\theta) \geq \epsilon_-/2$

$$\frac{a(\theta)w^{1-p_\Delta}(\theta)}{V^{\iota-p_\Delta\beta}(x)} = \frac{a(\theta)w(\theta)}{V^\iota(x)} \left(\frac{V^\beta(x)}{w(\theta)}\right)^{p_\Delta} \geq \frac{a(\theta)w(\theta)}{V^\iota(x)} (\epsilon_-/2)^{p_\Delta}$$

from which we conclude. We now partition $\mathcal{W}^c_{M_0} \times \mathsf{C}^c$ by considering the two following subsets.

(**i**) From our choice of M_0 and ϵ_- and (3.4) we deduce that on the subset $(\mathcal{W}^c_{M_0}\times\mathsf{C}^c)\cap\{\theta, x : V_\beta(\theta, x) < \epsilon_-\}$

$$\Delta(0) - \varepsilon - \Delta(V_\beta(\theta, x)) \leq -\varepsilon/2 \ ,$$

and consequently

$$\Lambda(\theta, x) \leq -\lambda V^\iota(x)/a(\theta) - w(\theta)\varepsilon/2 \ ,$$

and we conclude about the existence of λ_i,$\delta_i > 0$ such that (3.6) holds for any $\lambda \geq \lambda_i$ and $\tilde{\delta} = \delta_i$.

(**ii**) By our assumption on $\Delta(\cdot)$ there exists C'_{Δ,ϵ_-} such that for any $z \geq \epsilon_-$

$$\Delta(0) - \varepsilon - \Delta(z) \leq C'_{\Delta,\epsilon_-} z^{p_\Delta} \ .$$

Consequently we deduce that on $(\mathcal{W}^c_{M_0} \times \mathsf{C}^c) \cap \{\theta, x : V_\beta(\theta, x) \geq \epsilon_-\}$

$$\begin{aligned}
\Lambda(\theta, x) &\leq \tfrac{V^\iota(x)}{a(\theta)} \left[C'_{\Delta,\epsilon_-} \tfrac{a(\theta)w(\theta)}{V^\iota(x)} \left(\frac{V^\beta(x)}{w(\theta)} + c(\theta)\right)^{p_\Delta} - \lambda\right] \\
&\leq \tfrac{V^\iota(x)}{a(\theta)} \left[2^{p_\Delta} C'_{\Delta,\epsilon_-} \tfrac{a(\theta)w(\theta)^{1-p_\Delta}}{V^{\iota-p_\Delta\beta}(x)} - \lambda\right] \ .
\end{aligned}$$

We now choose $\lambda > 2^{p_\Delta} C'_{\Delta,\epsilon_-} C_{\epsilon_-}$, and from (3.7) with $q = 0$ we have $V^\iota(x)/a(\theta) \geq C^{-1}_{\epsilon_-} w(\theta)$ and therefore

$$\Lambda(\theta, x) \leq \frac{1}{2} \left[\tfrac{V^\iota(x)}{a(\theta)} + w(\theta)C^{-1}_{\epsilon_-}\right] \left(2^{p_\Delta} C'_{\Delta,\epsilon_-} C_{\epsilon_-} - \lambda\right) \ .$$

We conclude about the existence of λ_{ii},$\delta_{ii} > 0$ such that (3.6) holds for any $\lambda \geq \lambda_{ii}$ and $\tilde{\delta} = \delta_{ii}$.

(**c**) First, we note from our choice of M_0, (3.4) and (3.5) that for any $\theta, x \in \mathcal{W}^c_{M_0} \times \mathsf{C}$ and $\lambda > 0$ the function $\Lambda(\theta, x)$ is upper bounded by

$$\begin{aligned}
\Lambda(\theta, x) &\leq -\lambda V(x) + \lambda b(\theta) + [\Delta(0) - \varepsilon - \Delta(d(\theta))]\, w(\theta) \\
&\leq -\lambda V(x) + [\lambda b(\theta)/w(\theta) - \varepsilon/2]w(\theta) \quad .
\end{aligned}$$

We now show that for any $\lambda \in (0, +\infty)$ there exist $M_\lambda \in [M_0, +\infty)$ and $\delta_\lambda \in (0, +\infty)$ such that for all $\theta, x \in \mathcal{W}^c_{M_\lambda} \times \mathsf{C}$, (3.6) is satisfied with $\tilde{\delta} = \delta_\lambda$ a function of λ. From our last inequality and since $\iota \in [0, 1]$ and $V \geq 1$, for any $M \geq M_0$ and $\theta, x \in \mathcal{W}^c_M \times \mathsf{C}$

$$\Lambda(\theta, x) \leq -\lambda[\inf_{\vartheta\in\Theta} a(\vartheta)]\ V^\iota(x)/a(\theta) + [\lambda \sup_{\vartheta\in\mathcal{W}^c_M} b(\vartheta)/w(\vartheta) - \varepsilon/2]w(\theta) \ .$$

We conclude about the existence of M_λ and δ_λ as above from our assumption on $b(\cdot)$.

We now conclude by letting $\lambda_* \geq \lambda_{a,b} = \lambda_a \vee \lambda_b$,$M_* \geq M_{\lambda_a\vee\lambda_b}$ and $\delta = \delta_{\lambda_a\vee\lambda_b} \wedge \delta_a \wedge \delta_b$. □

4 The central role of stability in the context of stochastic approximation with Markovian dynamic

In this section we illustrate the central role played by the form of stability considered in this paper to establish that some controlled Markov chains of practical relevance possess some desired properties. We focus on a particular class of controlled Markov chains driven by a so-called stochastic approximation recursion (also known as the Robbins-Monro algorithm). The motivation for such algorithms, described below, is to find the roots of the function $h(\cdot) : \Theta \to \mathbb{R}^{n_\theta}$

$$h(\theta) := \int_{\mathsf{X}} H(\theta, x)\pi_\theta(dx) \ ,$$

for a family of functions $\{H(\theta, x) : \Theta \times \mathsf{X} \to \Theta\}$ and a family of probability distributions $\{\pi_\theta, \theta \in \Theta\}$ defined on some space $\mathsf{X} \times \mathcal{B}(\mathsf{X})$. This is a ubiquitous problem in statistics, engineering and computer science. These roots are rarely available analytically and a way of finding them numerically consists of considering the following controlled Markov chain on $\left((\Theta \times \mathsf{X})^{\mathbb{N}}, (\mathcal{B}(\Theta) \otimes \mathcal{B}(\mathsf{X}))^{\otimes \mathbb{N}}\right)$ initialised at some $(\theta_0, X_0) = (\theta, x) \in \Theta \times \mathsf{X}$ and defined recursively for a sequence of stepsizes $\{\gamma_i\}$ for $i \geq 0$,

$$\begin{aligned} X_{i+1} | \mathcal{F}_i &\sim P_{\theta_i}(X_i, \cdot) \\ \theta_{i+1} &= \theta_i + \gamma_{i+1} H(\theta_i, X_{i+1}) \end{aligned} \tag{4.1}$$

where $\{P_\theta, \theta \in \Theta\}$ (for some set $\Theta \subset \mathbb{R}^{n_\theta}$) is a family of Markov transition probabilities such that for each $\theta \in \Theta$, P_θ leaves π_θ invariant, *i.e.* is such that $\pi_\theta P_\theta = \pi_\theta$. The rational for this recursion is as follows. Let us first rewrite the Robbins-Monro recursion as

$$\theta_{i+1} = \theta_i + \gamma_{i+1} \left[h(\theta_i) + \xi_{i+1}\right]$$

where $\{\xi_{i+1} = H(\theta_i, X_{i+1}) - h(\theta_i)\}$, which is traditionally refered to as the "noise". Then $\{\theta_i\}$ can be thought as being a noisy version of the sequence $\{\bar{\theta}_i\}$ defined as $\bar{\theta}_{i+1} = \bar{\theta}_i + \gamma_{i+1} h(\bar{\theta}_i)$, and it is believable that the properties of $\{\theta_i\}$ are closely related to those of the noiseless sequence $\{\bar{\theta}_i\}$ provided the average effect of the noise on this sequence is negligible. This requires some form of averaging, or ergodicity, property on $\{\xi_i\}$.

The convergence of such sequences has been well studied by various authors, starting with the seminal work of [16], under various assumptions on all the quantities involved. A crucial step of such convergence analyses, however, consists of assuming that the sequence $\{\theta_i\}$ remains bounded in a compact set of Θ with probability one. This problem has traditionally been either ignored or circumvented by means of modifications of the recursion (4.1). Indeed, one of the major difficulties specific to the Markovian dynamic scenario is that $\{\theta_i\}$ governs the ergodicity of $\{X_i\}$ (and hence $\{\xi_i\}$) and that stability properties of $\{\theta_i\}$ relying on those of $\{\bar{\theta}_i\}$ require good ergodicity properties which might vanish whenever $\{\theta_i\}$ approaches a set $\partial\Theta$ away from the zeroes of $h(\theta)$, resulting in instability. Most existing results rely on modifications of the updates $\{\phi_\gamma\}$ designed to ensure a form of ergodicity of $\{\xi_i\}$ which in turn ensures that $\{\theta_i\}$ inherits the stability properties of $\{\bar{\theta}_i\}$. The only known results we are aware of where stability is established for (4.1) without any modification are [16, Part II, Section 1.9], where assumption (1.9.3) may not be satisfied in numerous cases of interest or directly verifiable, and [14] in a particular scenario.

The approach we follow here is significantly different from that developed in the aforementioned works and consists of dividing the difficult problem of proving boundedness away from $\partial\Theta$ into two simpler tasks. First using the results of Sections 2 or 3 one may establish that the sequence $\{\theta_i, X_i\}$ visits some set $\mathcal{W} \times C \subset \Theta \times \mathsf{X}$ infinitely often $\mathbb{P}_{\theta,x}$-a.s, a set which has the particularity that transition probabilities $\{P_\theta, \theta \in \mathcal{W}\}$ have uniformly good ergodicity properties. Then, using these facts one can show that $\{\theta_i\}$ follows the trajectories of the deterministic recursion $\{\bar{\theta}_i\}$ more and more accurately at each visit of $\mathcal{W} \times C$, and eventually remains in a set only slightly larger than $\mathcal{W}$ provided the deterministic sequence is itself stable. The advantage of our approach is that instead of aiming to establish ergodicity properties of $\{\xi_i\}$ in worse case scenarios for the sequence $\{\theta_i\}$ we decouple the analysis of the behaviour of $\{\theta_i\}$ when it approaches $\partial\Theta$ from the study of the ergodicity properties of $\{\xi_i\}$, which need to be studied for "reasonable" values of θ only. Before stating the main result of this section we state our assumptions.

(A4) Let $\{H(\theta, x)\}$, $\{\gamma_i\}$, $\{\pi_\theta\}$ and $\{P_\theta\}$ be as above. We assume that

1. there exists $\gamma_+ > 0$ such that,

(a) $\gamma := \{\gamma_i\} \subset [0, \gamma^+]^{\mathbb{N}}$,

(b) for any $\theta, x \in \Theta \times \mathsf{X}$ and $\gamma \in [0, \gamma^+]$

$$\theta + \gamma H(\theta, x) \in \Theta \ ,$$

2. $H : \Theta \times \mathsf{X} \to \mathbb{R}^{n_\theta}$ is such that for any $\theta \in \Theta$, $\int_{\mathsf{X}} |H(\theta, x)| \pi_\theta(dx) < +\infty$,
3. and for any $\theta \in \Theta$, $\pi_\theta P_\theta = \pi_\theta$.

A practical technique to prove the boundedness of the noiseless sequence consists, whenever possible, of determining a Lyapunov function $w : \Theta \to [0, \infty)$ such that $\langle \nabla w(\theta), h(\theta) \rangle \leq 0$ away from the roots of $h(\theta)$, where ∇w denotes the gradient of w with respect to θ and for $u, v \in \mathbb{R}^n$, $\langle \cdot, \cdot \rangle$ is their Euclidian inner product (we will later on also use the notation $|v| = \sqrt{\langle v, v \rangle}$ to denote the Euclidean norm of v). Note that although we use here the same symbol w as in Sections 1 or 3, the Lyapunov function below might be different.

(A5) Θ is an open subset of $\mathbb{R}^{n_\theta}$, $h : \Theta \to \mathbb{R}^{n_\theta}$ is continuous and there exists a continuously differentiable function $w : \Theta \to [0, \infty)$ such that,

1. there exists $M_0 > 0$ such that

$$\mathcal{L} := \{\theta \in \Theta, \langle \nabla w(\theta), h(\theta) \rangle = 0\} \subset \{\theta \in \Theta, w(\theta) < M_0\} \ ,$$

2. there exists $M_1 \in (M_0, \infty]$ such that $\mathcal{W}_{M_1}$ is a compact set,
3. for any $\theta \in \Theta \setminus \mathcal{L}$, $\langle \nabla w(\theta), h(\theta) \rangle < 0$.

We now introduce some additional notation needed to describe the ergodicity properties of $\{\xi_i\}$ every time the sequence $\{\theta_i, X_i\}$ visits some set $\mathcal{W} \times C$. More precisely, consider the stochastic processes $\{\vartheta_i, \mathfrak{X}_i\}$ defined on $\left((\Theta \times \mathsf{X})^{\mathbb{N}}, (\mathcal{B}(\Theta) \otimes \mathcal{B}(\mathsf{X}))^{\otimes \mathbb{N}}\right)$ for any $l \geq 0$, initialised with $\vartheta_0, \mathfrak{X}_0 \in \Theta \times \mathsf{X}$ and such that for $i \geq 0$ which uses the stepsize sequence $\gamma^{\leftarrow l} := \{\gamma_{i+l}, i \geq 0\}$

$$\begin{aligned}
\mathfrak{X}_{i+1} | (\vartheta_0, \mathfrak{X}_0, \mathfrak{X}_1, \ldots, \mathfrak{X}_i) &\sim P_{\vartheta_i}(\mathfrak{X}_i, \cdot) \\
\vartheta_{i+1} &= \vartheta_i + \gamma_{i+1+l} H(\vartheta_i, \mathfrak{X}_{i+1}) \ .
\end{aligned} \tag{4.2}$$

In order to take the shift in the stepsize sequence into account we denote by $\bar{\mathbb{P}}^{\gamma^{\leftarrow l}}_{\theta,x}$ and $\bar{\mathbb{E}}^{\gamma^{\leftarrow l}}_{\theta,x}$ the associated probability distribution and expectation operator for $\vartheta_0 = \theta \in \Theta$ and $\mathfrak{X}_0 = x \in \mathsf{X}$, and point out that in contrast to $\mathbb{P}_{\theta,x}$ defined earlier for $\{\theta_i, X_i\}$, the notational dependence on $\gamma^{\leftarrow l}$ for $l \geq 0$ is here crucial. For any $M > 0$ we define the exit time from the level set $\mathcal{W}_M$, $\sigma(\mathcal{W}_M) := \inf\{k \geq 1 : \vartheta_k \notin \mathcal{W}_M\}$ (with the standard convention that $\inf\{\emptyset\} = +\infty$) and for any $j \geq 1$ we define $\varsigma_j := H(\vartheta_{j-1}, \mathfrak{X}_j) - h(\vartheta_{j-1})$. We adapt the reasoning of [3, Proposition 5.2] (see also [2]) in order to establish the following result.

Theorem 3. *Assume (A4) and (A5), that $\{\gamma_i\}$ is such that $\limsup_{i\to\infty} \gamma_i = 0$ and let $M \in (M_0, M_1]$. Assume that there exists $\mathsf{C} \subset \mathsf{X}$ such that*

1. *for any $\epsilon > 0$,*

$$\limsup_{l\to\infty} \sup_{\theta, x \in \mathcal{W}_{M_0} \times \mathsf{C}} \bar{\mathbb{P}}^{\gamma^{\leftarrow l}}_{\theta,x} \left(\sup_{k \geq 1} \mathbb{I}\{\sigma(\mathcal{W}_M) \geq k\} \left| \sum_{j=1}^{k} \gamma_{j+l} \varsigma_j \right| > \epsilon \right) < 1 \ , \tag{4.3}$$

2. *for any $\theta, x \in \Theta \times \mathsf{X}$,*

$$\mathbb{P}_{\theta,x}\left(\cap_{k=1}^{\infty} \cup_{i=k}^{\infty} \{(\theta_i, X_i) \in \mathcal{W}_{M_0} \times \mathsf{C}\}\right) = 1 \ ,$$

i.e. $\{\theta_i, X_i\}$ defined by Eq. (4.1) visits $\mathcal{W}_{M_0} \times \mathsf{C}$ infinitely often $\mathbb{P}_{\theta,x}-$a.s.

Then the sequence $\{\theta_i\}$ as defined by Eq. (4.1) is such that

$$\mathbb{P}_{\theta,x}\left(\cup_{k=1}^{\infty} \cap_{i=k}^{\infty} \{\theta_i \in \mathcal{W}_M\}\right) = 1 \quad ,$$

that is $\{\theta_i\}$ eventually remains in $\mathcal{W}_M$, $\mathbb{P}_{\theta,x}-$a.s.

Remark 3. Proving Eq. (4.3) is now rather well understood in general scenarios as soon as some form of local (in θ) uniform ergodicity of $\{P_\theta\}$ is satisfied and can be checked in practice; see [2] and [3] for example and the recent results in [1]. In the present paper we rather focus on finding verifiable conditions on $\{\gamma_i\}, \{H(\theta,x)\}$ and $\{P_\theta\}$ which ensure that $\{\theta_i, X_i\}$ as defined by Eq. (4.1) visits $\mathcal{W}_{M_0} \times C$ infinitely often $\mathbb{P}_{\theta,x}-$a.s., which in combination with the aforementioned existing results will allow us to conclude about the stability of a large class of controlled MCMC algorithms.

Proof. For $M \in (M_0, M_1]$ we let $\delta_0 > 0$ and $\lambda_0 > 0$ be as in Theorem 7 from [3, Proposition 5.2] given below for convenience. We consider the sequence $\{T_i, i \geq 1\}$ of successive return times to $\mathcal{W}_{M_0} \times C$ "separated by at least an exit from $\mathcal{W}_M$", formally defined for $i \geq 0$ as

$$T_{i+1} = \inf\{j \geq T_i + 1 : \exists l \in \{T_i + 1, \ldots, j\}/\theta_l \notin \mathcal{W}_M \text{ and } (\theta_j, X_j) \in \mathcal{W}_{M_0} \times C\},$$

with the conventions $T_0 = 0$ and $\inf\{\emptyset\} = +\infty$. It will be useful below to note that for any $i \geq 1$, $T_i \geq i$. Let $n_0 \in \mathbb{N}$ be such that $\gamma_{T_{n_0}} \leq \gamma_{n_0} \leq \lambda_0$. We first show that for any $\theta, x \in \Theta \times \mathsf{X}$,

$$\mathbb{P}_{\theta,x}\left(\bigcup_{k\geq 1}\{T_k = +\infty\}\right) = 1 \ , \tag{4.4}$$

and to achieve this we establish a bound on $\sup_{\theta,x\in\Theta\times\mathsf{X}} \bar{\mathbb{P}}_{\theta,x}(T_n < +\infty)$ for $n \geq n_0$. Notice that from the strong Markov property, for any $\theta, x \in \Theta \times \mathsf{X}$ and $l \geq n_0$

$$\mathbb{P}_{\theta,x}(T_{l+1} < +\infty) = \mathbb{E}_{\theta,x}\left(\mathbb{I}\{T_l < +\infty\}\mathbb{P}_{\theta_{T_l},X_{T_l}}(T_1 < +\infty)\right) \ .$$

In addition,for any $\theta, x \in \Theta \times \mathsf{X}$ we have

$$\mathbb{I}\{T_l < +\infty\}\mathbb{P}_{\theta_{T_l},X_{T_l}}(T_1 < +\infty) \leq \mathbb{I}\{T_l < +\infty\}\bar{\mathbb{P}}^{\gamma^{\leftarrow T_l}}_{\theta_{T_l},X_{T_l}}(\sigma(\mathcal{W}_M) < +\infty) \quad \mathbb{P}_{\theta,x} - \text{a.s.}.$$

and for any $q \geq 0$

$$\bar{\mathbb{P}}^{\gamma^{\leftarrow q}}_{\theta,x}(\sigma(\mathcal{W}_M) < +\infty) = \bar{\mathbb{P}}^{\gamma^{\leftarrow q}}_{\theta,x}\left(\bigcup_{k\geq 1}\{\sigma(\mathcal{W}_M) = k\}\right) \ .$$

From Theorem 7 we deduce that for any $q \geq n_0$

$$\bigcup_{k\geq 1}\{\sigma(\mathcal{W}_M) = k\} \subset \left\{\sup_{k\geq 1} \mathbb{I}\{\sigma(\mathcal{W}_M) \geq k\}\left|\sum_{j=1}^{k} \gamma_{j+q}\varsigma_j\right| > \delta_0\right\} \ ,$$

which implies that for any $l \geq n_0$ (and hence $T_l \geq l \geq n_0$)

$$\mathbb{I}\{T_l < +\infty\}\mathbb{P}_{\theta_{T_l},X_{T_l}}(T_1 < +\infty) \leq \sup_{q\geq l} \sup_{\theta,x\in\mathcal{W}_{M_0}\times C} \bar{\mathbb{P}}^{\gamma^{\leftarrow q}}_{\theta,x}\left(\sup_{k\geq 1} \mathbb{I}\{\sigma(\mathcal{W}_M) \geq k\}\left|\sum_{j=1}^{k} \gamma_{j+q}\varsigma_j\right| > \delta_0\right) \ .$$

Consequently by induction one obtains that for any $n > n_0$,

$$\mathbb{P}_{\theta,x}(T_n < +\infty) \leq \prod_{l=n_0}^{n-1} \sup_{q\geq l} \sup_{\theta,x\in\mathcal{W}_{M_0}\times C} \bar{\mathbb{P}}^{\gamma^{\leftarrow q}}_{\theta,x}\left(\sup_{k\geq 1} \mathbb{I}\{\sigma(\mathcal{W}_M) \geq k\}\left|\sum_{j=1}^{k} \gamma_{j+q}\varsigma_j\right| > \delta_0\right) \ .$$

Result (4.4) then follows by a standard Borel-Cantelli argument under the condition of the theorem. We now prove that $\{\theta_k\}$ eventually remains in $\mathcal{W}_M$, $\mathbb{P}_{\theta,x}-$a.s.. First notice that by construction of $\{T_i\}$,

$$\bigcup_{k\geq 1}\{T_k = +\infty\} = \bigcup_{k\geq 1}\{T_{k-1} < +\infty, T_k = +\infty\} \ ,$$

and for any $k \geq 1$

$$\{T_{k-1} < +\infty, T_k = +\infty\} = \bigcup_{m\geq k-1}\{T_{k-1} = m, T_k = +\infty\}$$

and for any $m \geq k - 1$

$$\{T_{k-1} = m, T_k = +\infty\} = \{\theta_l \in \mathcal{W}_M, l \geq m + 1\} \cup \{\theta_l \notin \mathcal{W}_{M_0}, \ \forall l \geq m + 1\} \ .$$

As a result

$$\bigcup_{k\geq 1}\{T_k = +\infty\} = \bigcup_{m\geq 0}\{\theta_l \in \mathcal{W}_M, l \geq m + 1\} \cup \bigcup_{m\geq 0}\{\theta_l \notin \mathcal{W}_{M_0}, \ \forall l \geq m + 1\} \ .$$

Now, since by assumption

$$\mathbb{P}_{\theta,x}\left(\bigcup_{m\geq 1}\{\theta_l \notin \mathcal{W}_{M_0},\ \forall l \geq m+1\}\right) = 0$$

we conclude that for any $\theta, x \in \Theta \times \mathsf{X}$

$$\mathbb{P}_{\theta,x}\left(\bigcup_{m\geq 0}\{\theta_l \in \mathcal{W}_M, l \geq m+1\}\right) = 1 \ .$$

□

We briefly discuss here other applications of our stability results, in particular in the situation where the step-size sequence is held constant. Such fixed stepsize algorithms have been popular in engineering since they provide the algorithms with both some form of robustness and a "tracking" ability. The analysis of these algorithms naturally requires one to establish stability first [16]. We would like however to point out another important application in the context of adaptive step-size algorithms. Indeed, the choice of $\{\gamma_i\}$ is known to have an important impact on the convergence properties of $\{\theta_i\}$. In particular it is well known that if $\{\gamma_i\}$ vanishes too quickly in the early iterations of (4.1), convergence may be very slow. A natural way to address this problem consists of adaptively selecting the sequence of stepsizes $\{\gamma_i\}$. A strategy due to Kesten and further extended by Delyon and Juditsky in [7] is as follows. Given a non-increasing function $\gamma(\cdot) :\to (0,\infty)$ and $s_0 = 0$ consider the modification (ALG1) of 4.1.

Algorithm 1 Adaptive step-size algorithm

- $X_{i+1} \sim P_{\theta_i}(X_i, \cdot)$
- $\theta_{i+1} = \theta_i + \gamma(s_i) H(\theta_i, X_{i+1})$
- $s_{i+1} = s_i + \mathbb{I}\{\langle H(\theta_{i-1}, X_i), H(\theta_i, X_{i+1})\rangle < 0\}$

Assume for brevity that the root of $h(\theta) = 0$, θ_*, is unique. The rationale behind this recursion is that for sufficiently regular scenarios one may expect the event $\{\langle H(\theta_{i-1}, X_i), H(\theta_i, X_{i+1})\rangle < 0\}$ to occur with higher probability when θ_i is in a neighbourhood $B(\theta_*, \epsilon)$ of θ_* than when outside this neighbourhood. As a result $\gamma_i := \gamma(s_i)$ decreases slowly as long as $\{\theta_i\}$ is outside this neighbourhood of θ_*, and decreases much faster whenever $\{\theta_i\}$ aproaches θ^*. Convergence to θ_* requires that $\gamma_i \to 0$ or equivalently that $s_i \to \infty$ with probability one. This means that one should show that for any $\gamma \in \{\gamma(0), \gamma(1), \gamma(2), \ldots\}$ the fixed stepsize sequence $\vartheta_i^\gamma = \vartheta_{i-1}^\gamma + \gamma H(\vartheta_{i-1}^\gamma, X_i^\gamma)$ for $i \geq 1$ is recurrent in the aforementioned neighbourhood, which is the essence of the proof of [7]. Our results allow one to establish that the homogeneous Markov chain $\{\vartheta_i^\gamma, X_i^\gamma\}$ is recurrent in some set $\mathcal{W}_M \times \mathsf{C}$, the first crucial step of the proof of [7]. A detailed analysis of such a result is beyond the scope of the present paper.

5 Examples: some adaptive MCMC algorithms

In this section we illustrate how the results established in Sections 3 -4 can be straightforwardly applied to a variety of adaptive Markov chain Monte Carlo (MCMC) algorithms, where the aim is to automatically optimally choose the parameter θ of a family of Markov chain transition probabilities $\{P_\theta, \theta \in \Theta\}$, defined on some $\mathsf{X} \subset \mathbb{R}$ sharing a common invariant distribution with density $\pi(\cdot)$ with respect to the Lebesgue measure. More specifically, we focus here on the symmetric random walk Metropolis (SRWM) algorithm with transition probability defined for $(\theta, x, A) \in \Theta \times \mathsf{X} \times \mathcal{B}(\mathsf{X})$ as

$$P_\theta(x, A) = \int_{A-x} \alpha(x, x+z) q_\theta(z)\, dz + \mathbb{I}\{x \in A\} \int_{\mathsf{X}-x} (1 - \alpha(x, x+z))\, q_\theta(z)\, dz \quad . \tag{5.1}$$

where for any $x, y \in \mathsf{X}^2$, $\alpha(x,y) := 1 \wedge \pi(y)/\pi(x)$, and $\{q_\theta(\cdot), \theta \in \Theta\}$ is a family of symmetric increment probability densities with respect to the Lebesgue measure defined on $\mathsf{Z} \times \mathcal{B}(\mathsf{Z})$ for some $\mathsf{Z} \subset \mathsf{X}$. Various choices for $q_\theta(\cdot)$ are possible.

The AM (Adaptive Metropolis) algorithm of [10], is concerned with the situation where $\mathsf{X} = \mathbb{R}^{n_x}$ for some $n_x \geq 1$ and $\theta = [\mu|\Gamma] \in \Theta = \mathbb{R}^{n_x} \times \mathcal{C}$ where $\mathcal{C} \subset \mathbb{R}^{n_x \times n_x}$ is the cone of symmetric positive definite matrices and $q_\theta(z) := \det^{-1/2}((2.38^2/n_x)(\Gamma + \epsilon_{AM} I_{n_x \times n_x})) \times q\left(((2.38^2/n_x)(\Gamma + \epsilon_{AM} I_{n_x \times n_x}))^{-1/2} z)\right)$ for

$q(z) = \mathcal{N}(z; 0, I_{n_x \times n_x})$ and some $\epsilon_{AM} \in (0,1)$. In fact other choices for $q(\cdot)$ are possible as long as it is symmetric, that is $q(z) = q(-z)$ for all $z \in \mathsf{Z}$. In [8] it is shown that in some circumstances the "optimal" covariance matrix for the Normal-SRWM is Γ_π, where Γ_π is the true covariance matrix of the target distribution $\pi(\cdot)$, assumed here to exist. The AM algorithm of [10] essentially implements the following algorithm to estimate Γ on the fly. Let $\epsilon_{AM} > 0$ and let $X_0 = x \in \mathsf{X}$, then for $i \geq 0$ and with $\theta_i := [\mu_i | \Gamma_i]$ here,

Algorithm 2 AM algorithm, iteration $i+1$

- Sample $X_{i+1} \sim P_{\theta_i}(X_i, \cdot)$
- Update of the tuning parameter

$$\begin{aligned} \mu_{i+1} &= \mu_i + \gamma_{i+1}(X_{i+1} - \mu_i) \quad , \\ \Gamma_{i+1} &= \Gamma_i + \gamma_{i+1}((X_{i+1} - \mu_i)(X_{i+1} - \mu_i)^\mathsf{T} - \Gamma_i) \quad . \end{aligned} \tag{5.2}$$

It was realised in [4] that this algorithm is a particular instance of (4.1) where $H : \Theta \times \mathsf{X} \to \Theta$ is

$$H(\theta, x) := \left[x - \mu | (x - \mu)(x - \mu)^\mathsf{T} - \Gamma \right]^\mathsf{T} , \tag{5.3}$$

and the corresponding mean field is

$$h(\theta) = \left[\mu_\pi - \mu | (\mu_\pi - \mu)(\mu_\pi - \mu)^\mathsf{T} + \Gamma_\pi - \Gamma \right]^\mathsf{T} \quad .$$

We show in Subsection 5.2 (Theorem 5) that the stability of these recursions is a direct consequence of the result of Sections 3 and a result from [14], which establishes (A2) for a class of target distribution densities $\pi(\cdot)$. The result of Section4 then directly applies to the AM algorithm, leading to the conclusion that $\{\theta_i\}$ eventually remains in a compact set with probability one. While the boundedness of $\{\theta_i\}$ has already been established in [14] using different arguments our results are more general in several ways. For example Theorem 5 shows that the AM algorithm is stable when the sequence of stepsizes $\{\gamma_i\}$ is constant, which opens up the way for the analysis of more sophisticated and robust versions of the AM algorithm. Theorem 5 also shows that the AM algorithm is also stable for heavier tailed distributions than in [14], in the situation where $\mathsf{X} = \mathbb{R}$, for both decreasing or constant stepsize sequences. As should be clear from our current analysis, a full study of the multivariate scenario is a different (and significant) research project.

We now turn to another type of popular adaptive scheme for the SRWM. Let $\mathsf{X} = \mathbb{R}^{n_x}$ and $\Theta = \mathbb{R}$. Suppose $q(\cdot)$ is a symmetric probability density on X and define the family of proposal distributions $\{q_\theta(\cdot), \theta \in \Theta\}$ as $q_\theta(z) := \exp(-\theta) q(\exp(-\theta) z)$. Let $\alpha_* \in (0,1)$ be a desired mean acceptance probability for the SRWM. A possible increment probability density is again $q_\theta(z) = \mathcal{N}(z, \sigma := \exp(\theta))$. The following algorithm aims to optimise θ^* in order to achieve an expected acceptance rate of α_* [4] and is often used as one of the components of more sophisticated schemes.

Algorithm 3 Coerced acceptance probability RWM, iteration $i+1$

- Update the state X_i, Y_i, with $Z_{i+1} \sim q_{\theta_i}(\cdot)$

$$\begin{aligned} Y_{i+1} &= X_i + Z_{i+1} \\ X_{i+1} &= \begin{cases} Y_{i+1} & \text{with probability } \alpha(X_i, Y_{i+1}) \\ X_i & \text{otherwise} \end{cases} \quad , \end{aligned}$$

- Update the scaling parameter

$$\theta_{i+1} = \theta_i + \gamma_{i+1}(\alpha(X_i, Y_{i+1}) - \alpha_*) \quad .$$

In Subsection 5.1 we prove the stability of $\{\theta_i, X_i\}$ for a broad class of probability densities $\pi(\cdot)$, including a heavy tailed scenario and situations where the stepsize sequence is constant (Theorem 6). It should be pointed out that in this case, in contrast with the AM algorithm scenario, we do not require a lower bound

on the scaling factor $\exp(\theta)$, which requires establishing (A2) for both arbitrarily large and small values of $\exp(\theta)$ and leads us to proving the new result Theorem 4 (a stability result has been proved in [17], but in a less general scenario). In fact the theory we have developed suggests improvements on this standard algorithm whose stability can be easily established thanks to the theory developed earlier in the paper. An example is given below: the rationale behind the algorithm is that for very poor initialisations the increments on the parameter are initially large, while still leading to a stable dynamic.

Algorithm 4 Fast coerced acceptance probability RWM, iteration $i+1$

- Update the state X_i, Y_i, with $Z_{i+1} \sim q_{\theta_i}(\cdot)$

$$\begin{aligned} Y_{i+1} &= X_i + Z_{i+1} \\ X_{i+1} &= \begin{cases} Y_{i+1} & \text{with probability } \alpha(X_i, Y_{i+1}) \\ X_i & \text{otherwise} \end{cases} \end{aligned} \quad ,$$

- Update the scaling parameter

$$\theta_{i+1} = \theta_i + \gamma_{i+1}(|\theta_i| + 1)(\alpha(X_i, Y_{i+1}) - \alpha_*) \quad .$$

The proofs of stability of the three algorithms above rely on common key intermediate results. In Subsection 5.1 we establish (A2) for the SRWM under two different sets of assumptions on $\pi(\cdot)$ and $q(\cdot)$. In Subsection 5.2 we establish (A3) for the AM algorithm and conclude with Theorem 5, while in Subsection 5.3 we establish (A3) for the coerced acceptance algorithms, and conclude with Theorem 6.

5.1 Establishing (A2) for SRWM algorithms

5.1.1 The superexponential "$\Gamma + \epsilon_{AM} I_{n_x \times n_x}$" scenario

In this section we state a result which establishes that (A2) is satisfied for the SRWM transition probability on $\mathsf{X} = \mathbb{R}^{n_x}$ under suitable conditions on $\pi(\cdot)$ and $q(\cdot)$.

(**A6**) The probability distribution $\pi(\cdot)$ has the following properties:

1. it is positive on every compact set and continuously differentiable,
2. the exists $\rho > 1$ such that

$$\lim_{R\to+\infty} \sup_{\{x:|x|\geq R\}} \frac{x}{|x|^\rho} \cdot \nabla \log \pi(x) = -\infty \quad , \tag{5.4}$$

3. the contours $\partial A(x) = \{y : \pi(y) = \pi(x)\}$ are asymptotically regular, *i.e. for some* $R > 0$

$$\sup_{\{x:|x|\geq R\}} \frac{x}{|x|} \cdot \frac{\nabla \pi(x)}{|\nabla \pi(x)|} < 0 \quad , \tag{5.5}$$

4. the proposal distribution density q is that of a standardized Gaussian or Student's t-distribution.

The following theorem quantifies the way in which ergodicity of the SRWM vanishes under (A6) as some of its eigenvalues become large. The norm used for matrices below is $|A| = \sqrt{\mathsf{Tr}(AA^\mathsf{T})}$ and recall that here $\theta = [\mu|\Gamma] \in \Theta = \mathbb{R}^{n_x} \times \mathcal{C}$.

Proposition 2. *Let $\eta \in (0,1)$ and $V(x) \propto \pi^{-\eta}(x)$. Under (A6) one can choose $V \geq 1$ and there exist $a, b \in (0,\infty)$ and $\mathsf{C} = B(0,R)$ for some $R > 0$ such that for any $\theta, x \in \Theta \times \mathsf{X}$,*

$$P_\theta V(x) \leq (1 - a/\sqrt{\det(\Gamma)})V(x) + b\mathbb{I}\{x \in \mathsf{C}\} \ ,$$

and we note that for any $\Gamma \in \mathcal{C}$,

$$\sqrt{\det(\Gamma)} \leq n_x^{-n_x/4} |\Gamma|^{n_x/2} \quad .$$

Proof. The first statement is proved is in[14, Proposition 15], and the second statement follows from the standard arithmetic/geometric mean inequality applied to the eigenvalues of Γ^2,

$$\det(\Gamma^2) \leq \left(\frac{\mathsf{Tr}\,(\Gamma^2)}{n_x}\right)^{n_x} = \left(\frac{|\Gamma|^2}{n_x}\right)^{n_x} \quad . \tag{5.6}$$

□

5.1.2 Establishing (A2) for the AM algorithms with weak tail assumptions

In this section we prove (A2) for the SRWM algorithm on $\mathsf{X} = \mathbb{R}$ in the situation where no lower bound on the scaling parameter of the proposal distribution is assumed and under a weaker assumption on the vanishing rate of the tails of the target density $\pi(\cdot)$ than in the previous subsection. More precisely, let P_θ denote here the random-walk Metropolis kernel with symmetric proposal distribution $q_\theta(z) = \exp(-\theta)q(z/\exp(\theta))$ for $\theta \in \Theta := (-\infty, \infty)$. For notational simplicity, in this subsection, we introduce the piece of notation $\sigma = \exp(\theta)$ and use P_σ instead of $P_{\log\sigma}$, q_σ instead of $q_{\log\sigma}$ throughout and say that $\sigma \in \exp(\Theta)$. We will also use the piece of notation $\ell(x) := \log \pi(x)$. We require the following assumptions on $\pi(\cdot)$ and the increment proposal density q_σ

(A7)

1. The target distribution $\pi(\cdot)$ on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ has the following properties
   (a) It has a density $\pi(x)$ with respect to the Lebesgue measure,
   (b) $\pi(x)$ is bounded away from 0 on any compact set of $\mathbb{R}$,
   (c) $\ell(x)$ is twice differentiable. We denote $\ell'(x) := \nabla\ell(x)$ and $\ell''(x) := \nabla^2\ell(x)$,
   (d) for any $M > 0$, defining $\epsilon_x := M/|\ell'(x)|$,

$$\lim_{R\to\infty} \sup_{x\in B^c(0,R)} \sup_{|t|\leq\epsilon_x} \frac{|\ell''(x+t)|}{|\ell'(x+t)|^2} = 0 \quad ,$$

$$\lim_{R\to\infty} \sup_{x\in B^c(0,R)} \sup_{|t|\leq\epsilon_x} \frac{|\ell''(x+t)|}{|\ell'(x)|^2} = 0 \quad ,$$

$$\lim_{R\to\infty} \sup_{x\in B^c(0,R)} \sup_{0\leq t\leq\epsilon_x} \left|\frac{|\ell'(x-t)|}{|\ell'(x+t)|} - 1\right| = 0 \quad .$$

2. The tails of $\pi(x)$ decay at a minimum rate characterised as follows : there exist $\gamma, p \in (0,1)$ such that

$$\lim_{R\to\infty} \sup_{x\in B^c(0,R)} \frac{\ell'(x)}{|x|^{p-1}} < 0 \qquad \text{and} \qquad \lim_{R\to\infty} \inf_{x\in B^c(0,R)} \frac{\pi^{-\gamma}(x)}{|\ell'(x)|} > 0 \quad .$$

3. The increment proposal density $q_\sigma(z)$ is of the form $q_\sigma(z) = \frac{1}{\sigma}q(z/\sigma)$ for some symmetric probability density $q(z)$, such that $supp(q) = [-1,1] =: \mathsf{Z}$ and $q(\cdot) : \mathsf{Z} \to [\underline{q}, \bar{q}]$ for $\underline{q}, \bar{q} \in (0,\infty)$.

Remark 4. Consider $\ell(x) = C - |x|^\alpha$, for $|x| \geq R_\ell$ and $\alpha > 0$. Then for $x \geq R_\ell$, $\ell'(x) = -\alpha x^{\alpha-1}$ and $\ell''(x) = -\alpha(\alpha-1)x^{\alpha-2}$ and all the conditions in (A7) are satisfied.

Remark 5. The support condition on q can be removed but this requires one to control additional "tail integral" terms in the proofs of this section, which would add further to already long arguments. We have opted for this presentation for brevity and clarity since it is the terms that we handle which are both crucial and difficult to handle.

The following theorem establishes (A2) for the scalar SRWM with $V(x) \propto \pi^{-\beta}(x)$ for some $\beta \in (0,1)$ under (7).

Theorem 4. *Consider the SRWM targetting $\pi(\cdot)$ and with increment proposal density q_σ. Assume they satisfy (A7) and let $V(x) := c\pi^{-\eta}(x)$ for some $\eta \in (0,1)$ and $c \in (0,\infty)$ such that $V \geq 1$. Then for any $\iota \in (0,1)$ there exists $R \geq 0$ and $a_0 \in (0,\infty)$ such that*

1. for any $x \in B^c(0,R)$, with $a^{-1}(\sigma) := a_0/(\sigma \vee \sigma^{-2})$,

$$P_\sigma V(x) \leq V(x) - a^{-1}(\sigma)V^\iota(x) \quad ,$$

2. there exists a constant $b \in (0,\infty)$ such that for all $\sigma, x \in \exp(\Theta) \times B(0,R)$ we have $P_\sigma V(x) \le b$.

Proof. Let $\iota \in (0,1)$ and $R \ge R_0$ such that $\inf_{x\in B^c(0,R)} V^{1-\iota}(x)|\ell'(x)|^2 > 0$ and $\inf_{x\in B^c(0,R)} V^{1-\iota}(x)/|\ell'(x)| > 0$, where R_0 is given in Proposition 3. The existence of R is ensured by (A7)-2 and the choice of V. Indeed, from the assumption, for $x \in B^c(0,R)$ we have from Lemma 2 that $V(x) \ge C_{\Upsilon,2}\exp(\eta C_{\Upsilon,1}^{-1}|x|^p)$ for some constant $C > 0$ and $|\ell'(x)| \ge C_\ell |x|^{p-1}$ for $x \ge R_\ell$ for some $C_\ell, R_\ell > 0$, and we can conclude. From Proposition 3 below, for $x \in B^c(0,R)$ and $\sigma \le c_0/|\ell'(x)|$ we have

$$\begin{aligned}P_\sigma V(x) &\le V(x) - a_0'\sigma^2|\ell'(x)|^2V(x)\\ &\le V(x) - a_0' \inf_{x_0\in B^c(0,R)} |\ell'(x_0)|^2V^{1-\iota}(x_0) \times \sigma^2 V^\iota(x) \quad .\end{aligned}$$

For $|x| \ge \sigma \ge c_0/|\ell'(x)|$ we have

$$\begin{aligned}P_\sigma V(x) &\le V(x) - a_0' V(x)/|\sigma\ell'(x)|\\ &\le V(x) - a_0' \inf_{x_0\in B^c(0,R)} V^{1-\iota}(x_0)/|\ell'(x_0)|\sigma^{-1}V^\iota(x)\end{aligned}$$

For $\sigma \ge |x| \ge R$ we have

$$\begin{aligned}P_\sigma V(x) &\le V(x) - a_0'|x| \times V(x)/\sigma\\ &\le V(x) - a_0' R \times V(x)/\sigma\end{aligned}$$

Now we can use the trivial inequalities $\sigma^2 \le \sigma \le 1 \le \sigma^{-1}$ (case $\sigma \in (0,1]$) and $\sigma^{-1} \le 1 \le \sigma \le \sigma^2$ (case $\sigma \in (1,\infty)$) which lead to the following upper bound

$$P_\sigma V(x) \le V(x) - a_0(\sigma^{-1} \wedge \sigma^2)V^\iota(x)$$

The second claim follows immediately from the bound $PV(x) \le 2V(x)$ easily obtained from (5.9) and the fact that $\sup_{x\in B(0,R)} V(x) < \infty$. □

Proposition 3. *Consider the SRWM targetting $\pi(\cdot)$ satisfying (A7). Let $V(x) := c\pi^{-\eta}(x)$ for some $\eta \in (0,1)$ and c such that for all $x \in \mathsf{X}$, $V(x) \ge 1$. Then there exist $a_0', R_0 > 0$ such that for any $x \in B^c(0,R_0)$ and any $\sigma \in \exp(\Theta) = (0,\infty)$,*

$$\frac{P_\sigma V(x)}{V(x)} - 1 \le -a_0' \times \begin{cases} \sigma|\ell'(x)|^2, & \text{if } \sigma|\ell'(x)| < c_0\\ 1/|\sigma\ell'(x)|, & \text{if } |x| \ge \sigma \ge c_0/|\ell'(x)|\\ |x|/\sigma, & \text{if } 1 > |x|/\sigma\end{cases}$$

Proof. Without loss of generality we detail the situation where $x > 0$ since the case $x < 0$ can be straightforwardly addressed by considering the density $\pi_-(x) := \pi(-x)$ which also satisfies (A7), and hence Lemmata 3, 4 and 5. In what follows the terms $T_i(\sigma,x)$ for $i = 1,2,3,4$ are defined in Lemma 3. Choose $R \ge R_{PV} \vee R_\psi \vee R_T$ such that for $x \in B^c(0,R)$, $|x| \ge c_0/|\ell'(x)|$, where R_{PV}, R_ψ, R_T and c_0 are as in Lemmata 3, 4 and 5. First from Lemma 4, we have for $x \ge c_0/|\ell'(x)|$ and any $\sigma \in \exp(\Theta)$,

$$\begin{aligned}T_1(\sigma,x) &= \int_0^{\sigma\wedge c_0/|\ell'(x)|} \psi_x(z)q_\sigma(z)dz + \mathbb{I}\{\sigma \ge c_0/|\ell'(x)|\}\int_{c_0/|\ell'(x)|}^{\sigma\wedge x} \psi_x(z)q_\sigma(z)dz\\ &\le -\epsilon_\psi|\ell'(x)|^2\int_0^{\sigma\wedge c_0/|\ell'(x)|} z^2q_\sigma(z)dz - \epsilon_\psi\mathbb{I}\{\sigma \ge c_0/|\ell'(x)|\}\int_{c_0/|\ell'(x)|}^{\sigma\wedge x} q_\sigma(z)dz\\ &\le -\epsilon_\psi\underline{q}/3|\ell'(x)|^2[\sigma\wedge c_0/|\ell'(x)|]^3/\sigma - \epsilon_\psi\underline{q}\mathbb{I}\{\sigma \ge c_0/|\ell'(x)|\}[\sigma\wedge x - c_0/|\ell'(x)|]/\sigma \quad ,\end{aligned}$$

and therefore for $\sigma \le x$

$$\begin{aligned}T_1(\sigma,x) &\le -\epsilon_\psi\underline{q}\left[\mathbb{I}\{\sigma < c_0/|\ell'(x)|\}\sigma^2|\ell'(x)|^2/3 + \mathbb{I}\{\sigma \ge c_0/|\ell'(x)|\}\left(\frac{1}{3}c_0^3/|\ell'(x)|/\sigma + 1 - c_0/|\ell'(x)|/\sigma\right)\right]\\ &\le -\frac{1}{3}\epsilon_\psi\underline{q}\left[\sigma^2|\ell'(x)|^2 \times \mathbb{I}\{\sigma < c_0/|\ell'(x)|\} + c_0^3/|\ell'(x)|/\sigma \times \mathbb{I}\{x \ge \sigma \ge c_0/|\ell'(x)|\}\right] \quad .\end{aligned}$$

Now from Lemma 5 for $\sigma \geq x \geq R$ we have

$$T_1(\sigma,x)+T_2(\sigma,x) \leq -\epsilon_T \times x/\sigma$$
$$T_3(\sigma,x) \leq 0$$

and for $\sigma \geq x-\Upsilon(x)$ we have

$$T_3(\sigma,x)+T_4(\sigma,x) \leq -\epsilon_T \times (-\Upsilon(x))/\sigma$$

and we conclude with Lemma 3 and by treating the case where $x<0$ in a similar fashion. ☐

Note that, as pointed out in the proof of Proposition 3, it is sufficient to specialise most of the results of Lemmata 2, 3, 4 and 5 to the case $x>0$. The following lemma establishes some key properties implied specifically by (A7)-2, which will also be used in Subsection 5.3.

Lemma 2. *Assume that $\pi(\cdot)>0$, is differentiable and satisfies (A7)-2, define for any $\gamma>0$*

$$I_\gamma(x) := \int_0^\infty \left(\frac{\pi(x+\mathrm{sgn}(x)z)}{\pi(x)}\right)^\gamma dz \quad \text{and} \quad J_\gamma(x) := \int_0^{|x|} \left(\frac{\pi(x)}{\pi(x-\mathrm{sgn}(x)z)}\right)^\gamma dz \quad ,$$

with $\mathrm{sgn}(x):=x/|x|$ for $x\neq 0$ and for any $x>0$, $\Upsilon(x):=\inf\{y\in\mathsf{X}:\pi(y)=\pi(x)\}$. Then,

1. the function $\Upsilon(\cdot)$ has the following properties
   (a) $\lim_{x\to\infty}\Upsilon(x)=-\infty$,
   (b) there exist constants $C_{\Upsilon,1},C_{\Upsilon,2}\in(0,\infty)$ such that for all $|x|\geq R_\Upsilon$

   $$|\Upsilon(x)|\vee|x| \leq C_{\Upsilon,1}\left(-\log(\pi(x)/C_{\Upsilon,2})\right)^{1/p} \quad ,$$

   [or equivalently $\pi^{-1}(x)\geq C_{\Upsilon,2}\exp\left(C_{\Upsilon,1}^{-1}[|\Upsilon(x)|^p\vee|x|^p]\right)$].
2. and there exists a constant $C_\gamma\in(0,\infty)$ such that for any $x\in\mathsf{X}$, $I_\gamma(x)\vee J_\gamma(x)\leq C_\gamma|x|^{1-p}$.

Proof. First from assumption (A7)-2 there exist $R_\ell,C_\ell>0$ such that for all $x\in B^c(0,R_\ell)$ we have

$$\ell'(x) \leq -C_\ell|x|^{p-1} \quad ,$$

and consequently for all $x\in B^c(0,R_\ell)$ and $z\geq 0$ we have

$$\frac{\pi(x+\mathrm{sgn}(x)z)}{\pi(x)} = \exp\left(\mathrm{sgn}(x)\int_0^z \ell'(x+\mathrm{sgn}(x)t)dt\right) \leq \exp\left(-\frac{C_\ell}{p}\left[||x|+z|^p-|x|^p\right]\right)$$

Consequently for any $x\in B^c(0,R_\ell)$ we deduce that

$$\pi(x) \leq [\pi(-R_\ell)\vee\pi(R_\ell)]\exp(C_\ell/p|R_\ell|^p)\exp(-C_\ell/p|x|^p) \tag{5.7}$$

We deduce that there exists $R_1\geq R_\ell$ such that

$$[\pi(-R_\ell)\vee\pi(R_\ell)]\exp(C_\ell/p|R_\ell|^p)\exp(-C_\ell/pR_1^p) \leq \inf_{x\in B(0,R_\ell)}\pi(x) \quad ,$$

and from $\pi(\cdot)>0$, its continuity and the fact that it is monotone on both $(-\infty,-R_\ell]$ and $[R_1,\infty)$ we deduce the first statement. Now from (5.7) we deduce that there exists $C_1>0$ such that for $x\in[R_1,\infty)$

$$\pi(x)=\pi(\Upsilon(x)) \leq C_1\exp(-C_\ell|\Upsilon(x)|^p)$$

which implies the existence of $C_{\Upsilon,1},C_{\Upsilon,2},R_\Upsilon>0$ such that for any $x\in\mathsf{X}$ such that $|x|\geq R_\Upsilon$

$$|\Upsilon(x)|\vee|x| \leq C_{\Upsilon,1}\left(-\log(\pi(x)/C_{\Upsilon,2})\right)^{1/p} \tag{5.8}$$

From above we have the upper bound

$$I_\gamma(x) \leq \int_0^\infty \exp(-C_\ell\gamma/p\,[|x+\mathrm{sgn}(x)z|^p-|x|^p])dz$$

We can conclude with the result of Lemma 7. We proceed similarly with $J_\gamma(x)$ by noticing that $\ell(x)-\ell(x-\mathrm{sgn}(x)z)=\mathrm{sgn}(x)\int_{-z}^0\ell'(x+\mathrm{sgn}(x)t)dt\leq -C_\ell/p[|x|^p-|x-\mathrm{sgn}(x)z|^p]$ and again conclude with Lemma 7. ☐

We now find a convenient expression for $P_\sigma V(x)/V(x)$ valid for sufficiently large x and all σ's.

Lemma 3. *Assume (A7)-1 and for* $x,\eta,s,z \in \mathsf{X}\times(0,1)\times\{-1,1\}\times\mathsf{Z}$ *define* $\phi_{x,\eta,s}(z) := [\pi(x+sz)/\pi(x)]^\eta$ *and*

$$\psi_x(z) := \big(\phi_{x,-\eta,-1}(z)-1\big)+\big(\phi_{x,1-\eta,1}(z)-1\big)-\big(\phi_{x,1,1}(z)-1\big) \quad .$$

For any $x\geq 0$ *define* $\Upsilon(x) := \inf\{y\in\mathsf{X} : \pi(x)=\pi(y)\}$ *and let* $V(x)\propto\pi^{-\eta}(x)$*. Then there exists* $R_{PV}>0$ *such that for all* $x\geq R_{PV}$ *and any* $\sigma\in\exp(\Theta)$

$$\frac{P_\sigma V(x)}{V(x)}-1 \quad = \quad \sum_{i=1}^{4} T_i(\sigma,x)$$

with

$$T_1(\sigma,x) = \int_0^{\sigma\wedge x}\psi_x(z)q_\sigma(z)dz \quad ,$$

$$T_2(\sigma,x) = \mathbb{I}\{\sigma\geq x\}\int_x^{\sigma}\left[\phi_{x,1-\eta,1}(z)-\phi_{x,1,1}(z)\right]q_\sigma(z)dz \quad ,$$

$$T_3(\sigma,x) = \mathbb{I}\{\sigma\geq x\}\int_{\Upsilon(x)}^{(\sigma-x+\Upsilon(x))\wedge 0}\left[\phi_{\Upsilon(x),-\eta,-1}(z)-1\right]q_\sigma(z+x-\Upsilon(x))dz \quad ,$$

$$T_4(\sigma,x) = \mathbb{I}\{\sigma\geq x-\Upsilon(x)\}\int_0^{\sigma-(x-\Upsilon(x))}\left[\phi_{\Upsilon(x),1-\eta,-1}(z)-1+1-\phi_{\Upsilon(x),1,-1}(z)\right]q_\sigma(z+x-\Upsilon(x))dz \quad .$$

Proof. Let $\eta\in(0,1)$ and consider

$$\begin{aligned}
P_\sigma V(x) &= \int_{\mathsf{X}} V(y)\min\left\{1,\frac{\pi(y)}{\pi(x)}\right\}q_\sigma(x,y)\mathrm{d}y + V(x)\int_{\mathsf{X}}\left(1-\min\left\{1,\frac{\pi(y)}{\pi(x)}\right\}\right)q_\sigma(x,y)\mathrm{d}y\\
&= \int_{A_x} V(y)q_\sigma(x,y)\mathrm{d}y + \int_{R_x} V(y)\frac{\pi(y)}{\pi(x)}q_\sigma(x,y)\mathrm{d}y + V(x)\int_{R_x}\left(1-\frac{\pi(y)}{\pi(x)}\right)q_\sigma(x,y)\mathrm{d}y \quad ,
\end{aligned}$$

where $A_x := \{y\in\mathbb{R} : \pi(y)\geq\pi(x)\}$ and $R_x := \{y\in\mathbb{R} : \pi(y)<\pi(x)\}$ are the regions of (almost) sure acceptance and possible rejection, respectively. From this expression, we obtain

$$\begin{aligned}
\frac{P_\sigma V(x)}{V(x)}-1 &= \int_{A_x}\left(\frac{V(y)}{V(x)}-1\right)q_\sigma(x,y)\mathrm{d}y\\
&\quad + \int_{R_x}\left[\left(\frac{V(y)}{V(x)}\frac{\pi(y)}{\pi(x)}-1\right)+\left(1-\frac{\pi(y)}{\pi(x)}\right)\right]q_\sigma(x,y)\mathrm{d}y\\
&= \int_{A_x}\left[\left(\frac{\pi(y)}{\pi(x)}\right)^{-\eta}-1\right]q_\sigma(x,y)\mathrm{d}y\\
&\quad + \int_{R_x}\left\{\left[\left(\frac{\pi(y)}{\pi(x)}\right)^{1-\eta}-1\right]+\left[1-\frac{\pi(y)}{\pi(x)}\right]\right\}q_\sigma(x,y)\mathrm{d}y.
\end{aligned} \tag{5.9}$$

Notice that thanks to (A7) and Lemma 2 $\lim_{x\to\infty}\Upsilon(x)=-\infty$ and that for R sufficiently large, for any $x\geq R$ we have that $A_x=[\Upsilon(x),x]$ and $R_x=(-\infty,\Upsilon(x))\cup(x,\infty)$. Then with $y=x+z$ and by taking into account that the support of $q_\sigma(z)$ is included in $[-\sigma,\sigma]$ we have

$$\begin{aligned}
\frac{P_\sigma V(x)}{V(x)}-1 = &\int_0^{(x-\Upsilon(x))\wedge\sigma}\big(\phi_{x,-\eta,-1}(z)-1\big)q_\sigma(z)\mathrm{d}z + \int_0^{\sigma}\big(\phi_{x,1-\eta,1}(z)-1\big)-\big(\phi_{x,1,1}(z)-1\big)\big]q_\sigma(z)\mathrm{d}z\\
&+\mathbb{I}\{\sigma\geq x-\Upsilon(x)\}\int_{x-\Upsilon(x)}^{\sigma}\big(\phi_{x,1-\eta,-1}(z)-1\big)-\big(\phi_{x,1,-1}(z)-1\big)\big]q_\sigma(z)\mathrm{d}z
\end{aligned}$$

and therefore, because $x-\Upsilon(x)>x$, we may write

$$\begin{aligned}\frac{P_\sigma V(x)}{V(x)}-1=&\int_0^{\sigma\wedge x}\psi_x(z)q_\sigma(z)\mathrm{d}z+\mathbb{I}\{\sigma\geq x\}\int_x^\sigma\big(\phi_{x,1-\eta,1}(z)-1\big)-\big(\phi_{x,1,1}(z)-1\big)\big]q_\sigma(z)\mathrm{d}z\\&+\mathbb{I}\{\sigma\geq x\}\int_{x\wedge\sigma}^{(x-\Upsilon(x))\wedge\sigma}\big(\phi_{x,-\eta,-1}(z)-1\big)q_\sigma(z)\mathrm{d}z\\&+\mathbb{I}\{\sigma\geq x-\Upsilon(x)\}\int_{x-\Upsilon(x)}^{\sigma}\big(\phi_{x,1-\eta,-1}(z)-1\big)-\big(\phi_{x,1,-1}(z)-1\big)\big]q_\sigma(z)\mathrm{d}z\end{aligned}$$

and we conclude by using that $\pi(\Upsilon(x))=\pi(x)$ and the intermediate change of variable $z'=\Upsilon(x)-x+z$. $\square$

Here we prove some properties of $\psi_x(z)$ which will allow us to upper bound the term $T_1(\sigma,x)$ in the case where $\sigma\leq x$.

Lemma 4. *Assume (A7)-1 and for $\eta\in(0,1)$ let $\psi_x(z)$ be as in Lemma 3. Then there exist constants $c_0,\epsilon_\psi,R_\psi>0$ such that for all $x\geq R_\psi$, $\psi_x(\cdot):[0,x]\to(-\infty,0]$ and $\psi_x(z)$ satisfies the following upper bounds*

$$\psi_x(z)\leq-\epsilon_\psi\times\begin{cases}|\ell'(x)|^2z^2, & \text{for } 0\leq z\leq c_0/|\ell'(x)|\\ 1, & \text{for } c_0/|\ell'(x)|\leq z\leq x\end{cases}\tag{5.10}$$

Proof. Note first that for $s\in\{-1,1\}$, because $\phi_{x,\eta,s}(z):=[\pi(x+sz)/\pi(x)]^\eta=\exp[\eta(\ell(x+sz)-\ell(x))]$,

$$\begin{aligned}\phi'_{x,\eta,s}(z)&=\eta s\ell'(x+sz)\phi_{x,\eta,s}(z)\qquad\text{and}\\ \phi''_{x,\eta,s}(z)&=\big[\eta^2|\ell'(x+sz)|^2+\eta\ell''(x+sz)\big]\phi_{x,\eta,s}(z)\quad.\end{aligned}$$

We now prove the desired upper bounds on $\psi_x(z)$ by considering the following three cases (a) $0\leq z\leq c_0/|\ell'(x)|$, (b) $c_0/|\ell'(x)|\leq z\leq C_0/|\ell'(x)|$ and (c) $C_0/|\ell'(x)|\leq z\leq x$ for an appropriate choice of the constants $c_0,C_0>0$ to be determined.

Case (a) $0\leq z\leq c_0/|\ell'(x)|$. We consider a first-order Taylor expansion of $\psi_x(z)$ at $z_0=0$ with integral error form and obtain

$$\begin{aligned}\psi_x(z)&=z\eta\ell'(x)+z(1-\eta)\ell'(x)-z\ell'(x)+\int_0^z[\phi''_{x,-\eta,-1}(t)+\phi''_{x,1-\eta,1}(t)-\phi''_{x,1,1}(t)](z-t)\mathrm{d}t\\&=\int_0^z a_{x,\eta}(t)(z-t)\mathrm{d}t\end{aligned}$$

where

$$\begin{aligned}a_{x,\eta}(t):=&\,\eta^2|\ell'(x-t)|^2-\eta\ell''(x-t)\big]\phi_{x,-\eta,-1}(t)\\&+\big[(1-\eta)^2|\ell'(x+t)|^2+(1-\eta)\ell''(x+t)\big]\phi_{x,1-\eta,1}(t)-\big[|\ell'(x+t)|^2+\ell''(x+t)\big]\phi_{x,1,1}(t)\quad.\end{aligned}$$

We seek to upperbound $a_{x,\eta}(t)$. We choose $\epsilon_0\in(0,2\eta(1-\eta))$ and first show that for any $c_0\in(0,\epsilon_0/2)$

$$\lim_{x\to\infty}\inf_{0\leq z\leq c_0/|\ell'(x)|}\phi_{x,1,1}(z)\geq1-\epsilon_0/2\tag{5.11}$$

Indeed, for $0\leq z\leq c_0/|\ell'(x)|$ and x large enough to ensure $\ell'(x)<0$ we have for some $\xi_{x,z}\in[x,x+z]$ the following Taylor expansion

$$\begin{aligned}\ell(x+z)-\ell(x)&=\ell'(x)z+\frac{1}{2}z^2\ell''(x+\xi_{x,z})\\&\geq-c_0-\frac{c_0^2}{2}\frac{|\ell''(x+\xi_{x,z})|}{|\ell'(x)|^2}\end{aligned}$$

and with (A7)-1 the last term vanishes as $x\to\infty$ and we conclude by the assumption that $-c_0>-\epsilon_0/2$.

Now choose $\epsilon_1,\epsilon_2,\epsilon_3>0$. From (5.11) and (A7)-1 there exists $R>0$ such that for any $x\geq R$, $\inf_{|z|\leq c_0/|\ell'(x)|}\phi_{x,1,1}(z)\geq1-\epsilon_0/2$, $\sup_{|t|\leq c_0/|\ell'(x)|}|\ell''(x+t)|/|\ell'(x+t)|^2\leq\epsilon_1$, $\sup_{|t|\leq c_0/|\ell'(x)|}|\ell'(x-t)|^2/|\ell'(x+$

$t)|^2 \leq 1+\epsilon_2$ and $\sup_{|t|\leq c_0/|\ell'(x)|} |\ell''(x+t)|/|\ell'(x)|^2 \leq \epsilon_3/c_0$. With these, and observing that $0 \leq \phi_{\cdot,\cdot,\cdot}(t) \leq 1$, we obtain the following upper bound

$$a_{x,\eta}(t) \leq |\ell'(x)|^2 \frac{|\ell'(x+t)|^2}{|\ell'(x)|^2} \left[\eta(\eta+\epsilon_1)(1+\epsilon_2) + (1-\eta)(1-\eta+\epsilon_1) - (1-\epsilon_1)(1-\epsilon_0/2)\right] \quad .$$

We consider then the case where ϵ_0, $\epsilon_1$and ϵ_3 are chosen small enough so that the term in brackets in the last display is negative. We note now that since for some $\xi_{x,t} \in [x, x+t]$

$$\ell'(x+t) = \ell'(x) + t\ell''(x+\xi_{x,t})$$

then with $0 \leq t \leq c_0/|\ell'(x)|^2$ we have

$$\frac{\ell'(x+t)}{\ell'(x)} \geq 1 - c_0 \frac{|\ell''(x+\xi_{x,t})|}{|\ell'(x)|^2}$$

which leads to the following upperbound

$$a_{x,\eta}(t) \leq |\ell'(x)|^2(1-\epsilon_3)\left[\eta(\eta+\epsilon_1)(1+\epsilon_2) + (1-\eta)(1-\eta+\epsilon_1) - (1-\epsilon_1)(1-\epsilon_0/2)\right]$$

Notice that by our choice of ϵ_0 above, we have $\eta^2 + (1-\eta)^2 - (1-\epsilon_0/2) \leq -\eta(1-\eta)$. Now since $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ can be chosen arbitrarily small above we conclude about the existence of $M > 0$, $c_0 > 0$ and $R > 0$ such that for any $x \geq R$

$$\sup_{|t|\leq c_0/|\ell'(x)|} a_{x,\eta}(t) \leq -M|\ell'(x)|^2 \quad ,$$

and we therefore conclude that in such a case, for $0 \leq z \leq c_0/|\ell'(x)|$

$$\psi_x(z) \leq -M\frac{1}{2}z^2|\ell'(x)|^2 \quad .$$

Case (b) $c_0/|\ell'(x)| \leq z \leq C_0/|\ell'(x)|$. First notice that $\psi_x(0) = 0$ and inspect the derivative of this function and aim to prove that it is negative. For any $x \in \mathsf{X}$ we have

$$\begin{aligned}
\psi_x'(z) &= \eta\ell'(x-z)\phi_{x,-\eta,-1}(z) + (1-\eta)\ell'(x+z)\phi_{x,1-\eta,1}(z) - \ell'(x+z)\phi_{x,1,1}(z) \\
&= \ell'(x+z)\left[\eta\frac{\ell'(x-z)}{\ell'(x+z)}\phi_{x,-\eta,-1}(z) + (1-\eta)\phi_{x,1-\eta,1}(z) - \phi_{x,1,1}(z)\right] \quad .
\end{aligned}$$

Because $\ell'(x+z) < 0$ and the two first terms in brackets form a convex combination, the second line of (5.10) will be establish for $c_0/|\ell'(x)| \leq z \leq C_0/|\ell'(x)|$ once we will have shown that for $x \geq 0$ sufficiently large,

$$\phi_{x,1,1}(z) \leq \left(\frac{\ell'(x+z)}{\ell'(x-z)}\phi_{x,-\eta,-1}(z)\right) \wedge \phi_{x,1-\eta,1}(z) \quad .$$

Clearly $\phi_{x,1-\eta,1}(z) = \phi_{x,1,1}^{1-\eta}(z) \leq \phi_{x,1,1} \leq 1$, so we are left with showing that $\phi_{x,1,1}(z) \leq \frac{\ell'(x-z)}{\ell'(x+z)}\phi_{x,-\eta,-1}(z)$ or equivalently

$$\frac{\pi(x+z)}{\pi(x)}\left(\frac{\pi(x-z)}{\pi(x)}\right)^\eta \leq \frac{\ell'(x-z)}{\ell'(x+z)} \quad .$$

We consider the following Taylor expansion

$$\begin{aligned}
\ell(x+z) - \ell(x) + \eta\left[\ell(x-z) - \ell(x)\right] &= z\ell'(x) + \frac{1}{2}z^2\ell''(x+\xi_{x,z}) + \eta\left[-z\ell'(x) + \frac{1}{2}z^2\ell''(x+\xi_{x,-z})\right] \\
&= (1-\eta)z\ell'(x) + \frac{1}{2}z^2\left[\ell''(x+\xi_{x,z}) + \eta\ell''(x+\xi_{x,-z})\right]
\end{aligned}$$

for some $\xi_{x,z} \in [0, z]$ and $\xi_{x,-z} \in [-z, 0]$. For now choose any $C_0 > c_0$ and notice that for $c_0/|\ell'(x)| \leq z \leq C_0/|\ell'(x)|$ we have that

$$\begin{gathered}
(1-\eta)z\ell'(x) \leq -c_0(1-\eta) \\
z^2\left[|\ell''(x+\xi_{x,z})| + \eta|\ell''(x+\xi_{x,-z})|\right] \leq C_0^2\frac{|\ell''(x+\xi_{x,z})| + \eta|\ell''(x+\xi_{x,-z})|}{|\ell'(x)|^2} \quad .
\end{gathered}$$

Let $\epsilon_1 \in (0, c_0(1-\eta))$ and choose $\epsilon_2 > 0$ such that $\exp(-c_0(1-\eta)+\epsilon_1) < 1-\epsilon_2$. By (A7)-1 we can conclude by letting x be sufficiently large to ensure that for $c_0/|\ell'(x)| \le z \le C_0/|\ell'(x)|$,

$$\frac{\pi(x+z)}{\pi(x)}\left(\frac{\pi(x-z)}{\pi(x)}\right)^{\eta} \le \exp(-c_0(1-\eta)+\epsilon_1) < 1-\epsilon_2 \le \frac{\ell'(x-z)}{\ell'(x+z)} \quad .$$

Now using the result of case (a) we conclude that

$$\psi_x(z) \le \psi_x\left(\frac{c_0}{|\ell'(x)|}\right) \le -\frac{M}{2}c_0^2 \quad .$$

Case (c) $C_0/|\ell'(x)| \le z < x$. We have the following simple bound

$$\psi_x(z) \le \left(\frac{\pi(x)}{\pi(x-z)}\right)^{\eta} - 1 + \left(\frac{\pi(x+z)}{\pi(x)}\right)^{1-\eta} \quad . \tag{5.12}$$

We inspect, for $C_0/|\ell'(x)| \le z \le x$ and x large enough, the following difference

$$\begin{aligned}\ell(x+z) - \ell(x) &= \int_0^z \ell'(x+t)dt \\ &\le \int_0^{C_0/|\ell'(x)|} \ell'(x+t)dt \\ &\le -C_0 \sup_{0\le t\le C_0/|\ell'(x)|} \left|\frac{\ell'(x+t)}{\ell'(x)}\right| \quad ,\end{aligned}$$

and we can similarly obtain a bound on $\ell(x) - \ell(x-z) \le -C_0 \sup_{0\le t\le C_0/|\ell'(x)|}\left|\frac{\ell'(x-t)}{\ell'(x)}\right|$. From (A7)-1 and the Taylor expansion $\ell'(x+t) = \ell'(x) + z\ell''(x+\xi_{x,t})$ we conclude that for C_0 and x sufficiently large enough we can ensure that the upper bound in (5.12) is negative.

The proof is now concluded by choosing c_0 as in (a), which leads to the first line of (5.10), C_0 as in (c) and R large enough to cover cases (b) and (c), which imply the second line of (5.10). □

Now in this lemma we address the situation where $\sigma \ge x$ and require an additional assumption on the vanishing speed of $\ell'(x)$.

Lemma 5. *Assume (A7) and let $T_i(\sigma, x)$ for $i = 1, \ldots, 4$ be as defined in Lemma 3. Then there exist $C_T, R_T, \epsilon_T > 0$ such that for $x \ge R_T$ and*

1. for $\sigma \ge x$
$$T_1(\sigma,x) + T_2(\sigma,x) \le -\epsilon_T \times x/\sigma \quad ,$$
2. for $\sigma \ge x$
$$T_3(\sigma,x) \le 0 \quad ,$$
3. and for $\sigma \ge x - \Upsilon(x)$
$$T_3(\sigma,x) + T_4(\sigma,x) \le -\epsilon_T(-\Upsilon(x))/\sigma \quad .$$

Proof. We start with $T_1(\sigma,x) + T_2(\sigma,x)$ and with the notation of Lemma 2 we obtain

$$\begin{aligned}T_1(\sigma,x) + T_2(\sigma,x) \le \int_0^x \left[\left(\frac{\pi(x-z)}{\pi(x)}\right)^{-\eta} - 1\right] q_\sigma(z)dz + \int_0^\infty \left(\frac{\pi(x+z)}{\pi(x)}\right)^{1-\eta} q_\sigma(z)dz \\ \le \frac{q}{\sigma}\left[J_\eta(x) - x\right] + \frac{\bar{q}}{\sigma} I_{1-\eta}(x) \quad .\end{aligned}$$

For $\sigma \ge x$, because $\phi_{\Upsilon(x),1-\eta,1}(x) \le 1$ in the integration domain,

$$T_3(\sigma,x) \le 0 \quad .$$

For $\sigma \geq x - \Upsilon(x) \geq x$ we have on the one hand

$$\begin{aligned}
T_3(\sigma,x) &= \int_{\Upsilon(x)}^{0} \left[\phi_{\Upsilon(x),-\eta,-1}(z) - 1\right] q_\sigma(z + x - \Upsilon(x))dz \\
&\leq \frac{\underline{q}}{\sigma}\left(\Upsilon(x) + \int_{\Upsilon(x)}^{0} \phi_{\Upsilon(x),-\eta,-1}(z)dz\right) \\
&\leq \frac{\underline{q}}{\sigma}\left(\Upsilon(x) + \int_{0}^{-\Upsilon(x)} \phi_{\Upsilon(x),-\eta,1}(z)dz\right) \quad , \\
&\leq \frac{\underline{q}}{\sigma}\left(\Upsilon(x) + C|\Upsilon(x)|^{1-p}\right) \quad ,
\end{aligned}$$

where we have used Lemma 2. On the other hand we also have

$$\begin{aligned}
T_4(\sigma,x) &= \int_0^{\sigma-(x-\Upsilon(x))} \left[\left(\frac{\pi(\Upsilon(x)-z)}{\pi(\Upsilon(x))}\right)^{1-\eta} - \frac{\pi(\Upsilon(x)-z)}{\pi(\Upsilon(x))}\right] q_\sigma(z+x-\Upsilon(x))dz \quad , \\
&\leq \frac{\bar{q}}{\sigma}\int_0^\infty \left(\frac{\pi(\Upsilon(x)-z)}{\pi(\Upsilon(x))}\right)^{1-\eta} dz \\
&\leq \frac{C}{\sigma}(-\Upsilon(x))^{1-p} \quad ,
\end{aligned}$$

where we have used Lemma 2 again. We now conclude. □

5.2 Stability of the AM algorithms

Thanks to Theorem 1 and its corollary we know that recurrence is ensured as soon as (A2) and (A3) are satisfied. In the previous section we have established conditions on $\pi(\cdot)$ and $q_\theta(\cdot)$ under which (A2) is satisfied for the transition probabilities underpinning (ALG2). We therefore focus on checking that (A3) is satisfied. First we start with a result which together with Theorem 3 leads to the same conclusions as [14] when $\{\gamma_i\}$ is not constant, but also to the additional stability of the time-homogeneous Markov chain $\{\theta_i, X_i\}$ when $\gamma_i = \gamma_0$ for any $i \geq 0$.

Theorem 5. *Consider the controlled MC defined by (ALG2) for $\mathsf{X} = \mathbb{R}^{n_x}$ with $n_x \geq 1$ (resp. (ALG2) for $\mathsf{X} = \mathbb{R}$), assume that $\pi(\cdot)$ and $q(\cdot)$ satisfy (A6) (resp. (A7)) and that $\{\gamma_i\}$ is non-increasing and such that $\limsup_{i\to\infty} \gamma_{i+1}^{-1} - \gamma_i^{-1} < 1$. Then for any $\epsilon > 0$ there exist $M, R > 0$ such that with $\mathcal{W}_M := \{\theta \in \Theta : w(\theta) \leq M\}$ for $w(\theta) = 1 + |\mu|^{2+\epsilon} + |\Gamma|$*

$$\mathbb{P}_{\theta,x}\left(\cap_{k=0}^\infty \cup_{i\geq k} \{(\theta_i, X_i) \in \mathcal{W}_M \times B(0,R)\}\right) = 1 \quad .$$

The proofs of the theorem for the two sets of assumptions rely on the following proposition, which establishes (A3) for a suitable Lyapunov function $w(\cdot)$.

Proposition 4. *Let $\epsilon > 0$ and define $w : \Theta \to [0,\infty)$ as*

$$w(\theta) := 1 + |\mu|^{2+\epsilon} + |\Gamma|$$

and assume that (A2) holds for some $V : \mathsf{X} \to [1,\infty)$ such that for some $\beta \in (0,1)$ we have $V^\beta(x) \geq 1+|x|^{2+\epsilon}$ for all $x \in \mathsf{X}$. Let $\gamma^+ \in (0,1)$. Then there exists $C > 0$ such that for any $\gamma \in (0,\gamma^+]$, and any $\theta, x \in \Theta \times \mathsf{X}$

$$P_{\theta,\gamma} w(\theta,x) \leq w(\theta) - \gamma w(\theta)\Delta\left(w(\theta)^{-\epsilon/(2+\epsilon)} + \frac{V^\beta(x)\mathbb{I}\{x \notin \mathsf{C}\} + b^\beta(\theta)\mathbb{I}\{x \in \mathsf{C}\}}{w(\theta)}\right) \quad ,$$

where $b(\cdot) : \Theta \to [0,\infty)$ is as in (A2) and $\Delta(\cdot) : [0,\infty) \to \mathbb{R}$

$$\Delta(z) := 1 - C[z + z^{1/(2+\epsilon)}] \quad .$$

Proof. For $(x_+, \mu, \Gamma) \in \mathsf{X} \times \mathbb{R}^{n_x} \times \mathcal{C}$ and $\gamma \in [0,1]$, let $\mu_+ := \mu + \gamma[x_+ - \mu]$ and $\Gamma_+ := \Gamma + \gamma[(\mu - x_+)(\mu - x_+)^\mathsf{T} - \Gamma]$. We have the two trivial inequalities

$$\begin{aligned}
|\mu_+| &= |\mu + \gamma[x_+ - \mu]| \leq (1-\gamma)|\mu| + \gamma|x_+| \\
|\Gamma_+| &= |\Gamma + \gamma[(\mu - x_+)(\mu - x_+)^\mathsf{T} - \Gamma]| \leq (1-\gamma)|\Gamma| + \gamma|(\mu - x_+)(\mu - x_+)^\mathsf{T}|
\end{aligned}$$

which imply that with $w(\theta) := 1 + |\mu|^{2+\epsilon} + |\Gamma|$, denoting $\bar{\gamma} := \gamma/(1-\gamma) < 1/(1-\gamma^+)$,

$$
\begin{aligned}
w(\theta_+) - w(\theta) &\leq -|\mu|^{2+\epsilon} + \gamma\left[-|\Gamma| + |\mu - x_+|^2\right] + (1-\gamma)^{2+\epsilon}|\mu|^{2+\epsilon}\left[1 + \frac{\gamma}{1-\gamma}\frac{|x_+|}{|\mu|}\right]^{2+\epsilon} \\
&\leq \gamma\left[-|\Gamma| + |\mu - x_+|^2\right] + |\mu|^{2+\epsilon}\left[(1-\gamma)\left(1 + \bar{\gamma}|x_+|/|\mu|\right)^{2+\epsilon} - 1\right] \\
&\leq \gamma\left[-w(\theta) + 1 + |\mu|^{2+\epsilon} + 2\left(|\mu|^2 + |x_+|^2\right)\right] \\
&\quad + |\mu|^{2+\epsilon}\left[-\gamma\left(1 + \bar{\gamma}|x_+|/|\mu|\right)^{2+\epsilon} + \left(1 + \bar{\gamma}|x_+|/|\mu|\right)^{2+\epsilon} - 1\right]
\end{aligned}
$$

By the mean value theorem

$$
\begin{aligned}
|\mu|^{2+\epsilon}\left[\left(1 + \bar{\gamma}|x_+|/|\mu|\right)^{2+\epsilon} - 1\right] &\leq |\mu|^{2+\epsilon}(2+\epsilon)\bar{\gamma}|x_+|/|\mu| \times \left(1 + \bar{\gamma}|x_+|/|\mu|\right)^{1+\epsilon} \\
&= \gamma\frac{2+\epsilon}{1-\gamma} \times |\mu|^{1+\epsilon}|x_+|\left(1 + \bar{\gamma}|x_+|/|\mu|\right)^{1+\epsilon}
\end{aligned}
$$

and since $|\mu|^{2+\epsilon}[1 - (1 + \bar{\gamma}|x_+|/|\mu|)^{2+\epsilon}] \leq 0$ we obtain the following bound,

$$
\begin{aligned}
w(\theta_+) - w(\theta) \leq \gamma w(\theta)\left[-1 + \frac{1}{w(\theta)} + \frac{2|\mu|^2}{w(\theta)}\left(1 + |x_+|^2/|\mu|^2\right) + \frac{(2+\epsilon)}{1-\gamma}\frac{|\mu|^{1+\epsilon}|x_+|}{w(\theta)}\left(1 + \bar{\gamma}|x_+|/|\mu|\right)^{1+\epsilon}\right] \\
\leq \gamma w(\theta)\left(-1 + C\Psi(\theta, x_+)\right) \quad ,
\end{aligned}
$$

for some $C \in (0, \infty)$ and where

$$
\Psi(\theta, x_+) := \left(\frac{|\mu|^2}{w(\theta)} + \frac{1 + |x_+|^2}{w(\theta)}\right) + \left(\frac{|\mu| \times |x_+|^{1/(1+\epsilon)}}{w^{1/(1+\epsilon)}(\theta)} + \frac{|x_+|^{1+1/(1+\epsilon)}}{w^{1/(1+\epsilon)}(\theta)}\right)^{1+\epsilon} \quad .
$$

Now the identity $(a+b)^{1+\epsilon} \leq 2^\epsilon(a^{1+\epsilon} + b^{1+\epsilon})$ for $a, b > 0$ and the following equalities

$$
\begin{aligned}
\frac{|\mu| \times |x_+|^{1/(1+\epsilon)}}{w^{1/(1+\epsilon)}(\theta)} &= \frac{|\mu|}{w^{1/(2+\epsilon)}(\theta)}\left(\frac{|x_+|^{2+\epsilon}}{w(\theta)}\right)^{1/(1+\epsilon)/(2+\epsilon)} \\
\frac{|x_+|^{1+1/(1+\epsilon)}}{w^{1/(1+\epsilon)}(\theta)} &= \left(\frac{|x_+|^{2+\epsilon}}{w(\theta)}\right)^{1/(1+\epsilon)}
\end{aligned}
$$

yield

$$
\left(\frac{|\mu| \times |x_+|^{1/(1+\epsilon)}}{w^{1/(1+\epsilon)}(\theta)} + \frac{|x_+|^{1+1/(1+\epsilon)}}{w^{1/(1+\epsilon)}(\theta)}\right)^{1+\epsilon} \leq 2^\epsilon\left(\left(\frac{|\mu|^{2+\epsilon}}{w(\theta)}\right)^{\frac{1+\epsilon}{2+\epsilon}}\left(\frac{|x_+|^{2+\epsilon}}{w(\theta)}\right)^{1/(2+\epsilon)} + \frac{|x_+|^{2+\epsilon}}{w(\theta)}\right)
$$

and

$$
\frac{|\mu|^2}{w(\theta)} = \frac{\left(|\mu|^{2+\epsilon}\right)^{2/(2+\epsilon)}}{w(\theta)} \leq w(\theta)^{-\epsilon/(2+\epsilon)} \quad .
$$

We therefore deduce that if for any $x \in \mathsf{X}$, $V^\beta(x) \geq 1 + |x|^{2+\epsilon}$ then

$$
\begin{aligned}
\Psi(\theta, x_+) &\leq w(\theta)^{-\epsilon/(2+\epsilon)} + \frac{1 + |x_+|^2}{w(\theta)} + 2^\epsilon\left(\left(\frac{|x_+|^{2+\epsilon}}{w(\theta)}\right)^{1/(2+\epsilon)} + \frac{|x_+|^{2+\epsilon}}{w(\theta)}\right) \\
&\leq w(\theta)^{-\epsilon/(2+\epsilon)} + 2\frac{V^\beta(x_+)}{w(\theta)} + 2^\epsilon\left(\left(\frac{V^\beta(x_+)}{w(\theta)}\right)^{1/(2+\epsilon)} + \frac{V^\beta(x_+)}{w(\theta)}\right) \quad .
\end{aligned}
$$

Now by (A2) and Jensen's inequality we deduce that for some constant $C > 1$

$$
\begin{aligned}
&P_{\gamma,\theta}\Psi(\theta, x) \\
&\quad \leq w(\theta)^{-\epsilon/(2+\epsilon)} + C\left[\left(\frac{V^\beta(x)\mathbb{I}\{x \notin \mathsf{C}\} + b^\beta(\theta)\mathbb{I}\{x \in \mathsf{C}\}}{w(\theta)}\right)^{1/(2+\epsilon)} + \frac{V^\beta(x)\mathbb{I}\{x \notin \mathsf{C}\} + b^\beta(\theta)\mathbb{I}\{x \in \mathsf{C}\}}{w(\theta)}\right]
\end{aligned}
$$

and conclude. □

5.2.1 Multivariate case and superexponential tails: (A6)

Proof of Theorem 5 under (A6). Let $\epsilon > 0$, $\beta \in (0, 1/(1 + n_x/2)]$ and $V(x) \propto \pi^{-\eta}(x)$ for some $\eta \in (0,1)$ where the constant of proportionality is such that $V^{\beta}(x) \geq 1 + |x|^{2+\epsilon}$ (which is possible as $\pi^{-\eta}(x) \geq C_1 \exp(C_2|x|)$ for some $C_1, C_2 > 0$). From Proposition 2 there exists $a', b' > 0$ and $\mathsf{C} := B(0, R)$ for some $R > 0$ such that for any $x \in \mathsf{X}$

$$\begin{aligned} P_\theta V(x) &\leq (1 - a'/|\Gamma + \epsilon_{AM} I_{n_x \times n_x}|^{n_x/2}) V(x) + b' \mathbb{I}\{x \in \mathsf{C}\} \\ &\leq \left[1 - a''/\left(|\epsilon_{AM} I_{n_x \times n_x}|^{n_x/2} + w^{n_x/2}(\theta)\right)\right] V(x) \mathbb{I}\{x \notin \mathsf{C}\} + b \mathbb{I}\{x \in \mathsf{C}\} \ , \end{aligned}$$

where $b = b' + \sup_{x \in \mathsf{C}} V(x)$. Now from Proposition 4, with $w(\theta) := 1 + |\mu|^{2+\epsilon} + |\Gamma|$, we have

$$P_{\theta,\gamma} w(\theta, x) \leq w(\theta) - \gamma w(\theta) \Delta \left(w(\theta)^{-\epsilon/(2+\epsilon)} + \frac{V^{\beta}(x) \mathbb{I}\{x \notin \mathsf{C}\} + b^{\beta} \mathbb{I}\{x \in \mathsf{C}\}}{w(\theta)} \right)$$

with $\Delta(z) := 1 - C[z + z^{1/(2+\epsilon)}]$ and we conclude with Proposition 1 and Theorem 2 . □

5.2.2 Relaxed tail conditions, univariate scenario: (A7)

Now we draw the same conclusions when $\mathsf{X} = \mathbb{R}$ and $\pi(\cdot)$ now satisfies less stringent tail conditions.

Proof of Theorem 5 under (A7). Let $\iota, \eta \in (0,1)$ and $\beta \in (0, \iota/2]$. Let $V(x) \propto \pi^{-\eta}(x)$, such that $V^{\beta}(x) \geq 1 + |x|^{2+\epsilon}$ (which is possible as $\pi^{-\eta}(x) \geq C_1 \exp(C_2|x|^p)$ for some $C_1, C_2 > 0$ from Lemma 2). From Theorem 4, there exist $b, R > 0$ such that with $\mathsf{C} = B(0, R)$ for any $\theta, x \in \Theta \times \mathsf{X}$

$$P_\theta V(x) \leq \left[V(x) - a^{-1}(\theta) V^{\iota}(x)\right] \mathbb{I}\{x \in \mathsf{C}\} + b \mathbb{I}\{x \in \mathsf{C}\} \quad ,$$

with $a^{-1}(\theta) = a_0/[(\Gamma + \epsilon_{AM})^{-2} \vee (\Gamma + \epsilon_{AM})] \geq a_0/[(\epsilon_{AM}^{-2} \vee (\epsilon_{AM} + w(\theta))]$ and $w(\theta) := 1 + |\mu|^{2+\epsilon} + |\Gamma|$. From Proposition 4 we therefore have

$$P_{\theta,\gamma} w(\theta, x) \leq w(\theta) - \gamma w(\theta) \Delta \left(w(\theta)^{-\epsilon/(2+\epsilon)} + \frac{V^{\beta}(x) \mathbb{I}\{x \notin \mathsf{C}\} + b^{\beta} \mathbb{I}\{x \in \mathsf{C}\}}{w(\theta)} \right)$$

with $\Delta(z) := 1 - C[z + z^{1/(2+\epsilon)}]$ and we conclude with Proposition 1 and Theorem 2 . □

5.3 Stability of the coerced acceptance probability algorithms

In this subsection we establish the stability of (ALG3) and (ALG4) in a univariate setting. We proceed as in Section 5.2 and aim to apply Theorem 1 and its corollary which require (A2) and (A3) to be satisfied. A related result has been established in [17] under a more stringent condition on the decay of the tails of the target density, and not covering constant stepsize sequences $\{\gamma_i\}$.

Theorem 6. *Consider the controlled MC as defined by either (ALG3) or (ALG4) for some $\alpha_* \in (0, 1/2)$. Assume that $\pi(\cdot)$ and $q(\cdot)$ satisfy (A7) and that the stepsize sequence $\{\gamma_i\}$ is such that*

$$(\limsup_{i\to\infty} \gamma_{i+1}^{-1} - \gamma_i^{-1}) + \limsup_{i\to\infty} \gamma_i < \alpha_* \wedge (\frac{1}{2} - \alpha_*) \quad .$$

Let $w(\theta) = \exp(|\theta|)$ for (ALG3) and $w(\theta) := 1 + |\theta|^2$ for (ALG4). Then there exist $M, R > 0$ such that for any $\theta, x \in \Theta \times \mathsf{X}$,

$$\mathbb{P}_{\theta,x} \left(\cap_{k=0}^{\infty} \cup_{i \geq k} \{(\theta_i, X_i) \in \mathcal{W}_M : \times B(0, R)\} \right) = 1 \quad ,$$

i.e. $\mathbb{P}_{\theta,x} - a.s.$ $\{\theta_i, X_i\}$ visits $\mathcal{W}_M \times B(0, R)$ infinitely often.

5.3.1 Proof in the standard scenario: (ALG3)

Before starting the proofs it is worth stressing on the fact that throughout

$$P_\theta(x, y; \mathrm{d}x' \times \mathrm{d}y') = q(x, \mathrm{d}y')[\alpha(x, y')\delta_{y'}(\mathrm{d}x') + (1 - \alpha(x, y'))\, \delta_x(\mathrm{d}x')]$$

and hence that for any $x, y \in \mathsf{X}$, $P_\theta(x, y; \cdot) = P_\theta(x, \cdot)$ and that for notational simplicity the Lyapunov function $V(x)$ should be understood as being the function $V(x) \times 1$ defined on X^2.

(Proof of Theorem 6 in the case of (ALG3)). First notice that there exists $i_0 \in \mathbb{N}$ such that $\sup_{i \geq i_0} \gamma_{i+1}^{-1} - \gamma_i^{-1} < \alpha_* \wedge (\frac{1}{2} - \alpha_*) - \gamma_{\max}$ with $\gamma_{\max} := \sup_{i \geq i_0} \gamma_i$. We show that (A2) and (A3) are satisfied and conclude with Theorem 2 and Corollary 1, for $i \geq i_0$. Let $\eta, \iota \in (0,1)$, $\beta \in (0, \iota/3]$ and define $V(x) \propto \pi^{-\eta}(x)$, such that $V^\beta(x) \geq 1 \vee (-\log \pi(x))^{1/p}$ (which is possible from Lemma 2). From Theorem 4, there exist $b, R > 0$ such that for any $\theta, x \in \Theta \times \mathsf{X}$

$$P_\theta V(x) \leq \left[V(x) - a^{-1}(\theta) V^\iota(x)\right] \mathbb{I}\{x \notin \mathsf{C}\} + b\mathbb{I}\{x \in \mathsf{C}\} \quad ,$$

with $a(\theta) = [\exp(\theta) \vee \exp(-2\theta)]/a_0$ (for some $a_0 > 0$) and $\mathsf{C} = B(0, R)$. Now with $w(\theta) = \exp(|\theta|)$ from Lemma 6 there exists $C > 0$ such that

$$\begin{aligned}
&\alpha_{\exp(\theta)}(x) \geq 1/2 - C\frac{V^\beta(x)}{\exp(-\theta)} && \text{for } \theta \leq 0 \text{ and } x \in \mathsf{X} \\
&\alpha_{\exp(\theta)}(x) \leq C_+ \frac{(-\log \pi(x))^{1/p} \vee 1}{\exp(\theta)} \leq C\frac{V^\beta(x)}{\exp(\theta)} && \text{for } \theta \geq 0 \text{ and } x \in \mathsf{X}.
\end{aligned}$$

One can apply Proposition 5, leading to the existence of $C > 0$ such that for any $\theta, x \in \Theta \times \mathsf{X}$,

$$P_{\gamma,\theta} w(\theta, x) \leq w(\theta) - \gamma w(\theta) \Delta\left(\left[c(\theta) + \frac{V^\beta(x)}{w(\theta)}\right] \mathbb{I}\{x \notin \mathsf{C}\} + d(\theta)\mathbb{I}\{x \in \mathsf{C}\}\right) \quad ,$$

with

$$c(\theta) = C^{-1}\mathbb{I}\{|\theta| \leq \gamma_{\max}\}(2 + \alpha_* \wedge (\frac{1}{2} - \alpha_*)) \quad ,$$

$$d(\theta) = \frac{\sup_{x \in \mathsf{C}} V^\beta(x)}{w(\theta)} + C^{-1}\mathbb{I}\{|\theta| \leq \gamma_{\max}\}(2 + \alpha_* \wedge (\frac{1}{2} - \alpha_*)) \quad ,$$

and

$$\Delta\left(z\right) = \alpha_* \wedge (\frac{1}{2} - \alpha_*) - \gamma_{\max} - Cz \quad .$$

Notice that from our choice of i_0 $\sup_{i \geq i_0} \gamma_{i+1}^{-1} - \gamma_i^{-1} < \Delta(0)$ and that we have $\iota/\beta \geq 3 > p_\Delta = 1$ in (A3). Now since $\beta \leq \iota/3$ and $a(\theta) \leq w^2(\theta)/a_0$ we use Proposition 1 and conclude. $\square$

Proposition 5. *Consider the controlled MC as defined in (ALG3) with $\alpha_* \in (0, 1/2)$, $V(x) := c\pi^{-\eta}(x)$ and assume that there exist $C > 0$ and $\beta \in [0, 1)$ such that for all $(\theta, x) \in \Theta \times \mathsf{X}$*

$$\operatorname{sgn}(\theta)(\alpha_{\exp(\theta)}(x) - \alpha_*) \leq -[\alpha_* \wedge (1/2 - \alpha_*)] + CV^\beta(x)/\exp(|\theta|) \quad .$$

Let $\gamma_{\max} \in (0, \alpha_* \wedge (1/2 - \alpha_*))$. Then for any $\gamma \in (0, \gamma_{\max}]$ and $\theta, x \in \Theta \times \mathsf{X}$ and the Lyapunov function $w : \Theta \to [1, \infty)$ defined by $w(\theta) := \exp|\theta|$

$$P_{\gamma,\theta} w(\theta, x) \leq w(\theta) - \gamma w(\theta) \Delta\left(\mathbb{I}\{|\theta| \leq \gamma_{\max}\} C^{-1}(2 + \alpha_* \wedge (\frac{1}{2} - \alpha_*)) + \frac{V^\beta(x)}{w(\theta)}\right) \quad ,$$

with

$$\Delta\left(z\right) = \alpha_* \wedge (\frac{1}{2} - \alpha_*) - \gamma_{\max} - Cz \quad .$$

Proof. For $|\theta| > \gamma$ and since for all $x, y_+ \in \mathsf{X}^2$ $|\alpha(x, y_+) - \alpha_*| \leq 1$, one can write (with $\theta_+ = \theta + \gamma[\alpha(x, y_+) - \alpha_*]$)

$$\begin{aligned}
w(\theta_+) &= w(\theta + \gamma[\alpha(x, y_+) - \alpha_*]) \\
&= \exp(|\theta| + \operatorname{sgn}(\theta)\gamma[\alpha(x, y_+) - \alpha_*]) \quad .
\end{aligned}$$

Now since $\gamma \leq 1$, from the inequality $\exp(u) \leq 1 + u + u^2$ valid for $|u| \leq 1$, one obtains

$$w(\theta_+) \leq w(\theta)(1 + \gamma \operatorname{sgn}(\theta)[\alpha(x, y_+) - \alpha_*] + \gamma^2) \quad .$$

Taking the conditional expectations yields for $|\theta| > \gamma$

$$\begin{aligned}
P_{\gamma,\theta} w(\theta, x) &\leq w(\theta)(1 + \gamma \operatorname{sgn}(\theta)[\alpha_{\exp(\theta)}(x) - \alpha_*] + \gamma^2) \quad . \\
&\leq w(\theta) - \gamma w(\theta)\left([\alpha_* \wedge (1/2 - \alpha_*)] - CV^\beta(x)/\exp(|\theta|) - \gamma_{\max}\right) \quad .
\end{aligned}$$

Also notice that for any $\theta \in \Theta$ we have $|\theta_+| \leq |\theta| + \gamma$, whence $w(\theta_+) \leq w(\theta)\exp(\gamma) \leq w(\theta)(1+\gamma+\gamma^2) \leq w(\theta)(1+2\gamma)$ whenever $\gamma \leq 1$. From this inequality and the display above we deduce for all $\theta, x \in \Theta \times \mathsf{X}$ and $\gamma \in (0, \gamma_{\max}]$,

$$P_{\gamma,\theta}w(\theta,x) \leq w(\theta) - \gamma w(\theta)\left[\alpha_* \wedge (\frac{1}{2} - \alpha_*) - \gamma_{\max} - \mathbb{I}\{|\theta| \leq \gamma_{\max}\}(2 + \alpha_* \wedge (\frac{1}{2} - \alpha_*)) - CV^\beta(x)/w(\theta)\right] \quad .$$

□

Lemma 6. *Assume that $\pi(\cdot)$ is a strictly positive, differentiable probability density satisfying (A7)-2. Moreover, suppose that $q_\sigma(z) := \sigma^{-1}q(z/\sigma)$ where $q : \mathsf{X} \to [0, \bar{q}]$ for $\bar{q} > 0$ and such that it has a finite first order moment. Then, there exists constants $C_-, C_+ > 0$ such that for any $\alpha_* \in (0, 1/2)$*

$$\alpha_\sigma(x) \geq 1/2 - C_-\sigma \qquad \text{for } \sigma \leq 1 \text{ and } x \in \mathsf{X}$$

$$\alpha_\sigma(x) \leq C_+\frac{(-\log\pi(x))^{1/p} \vee 1}{\sigma} \qquad \text{for } \sigma \geq 1 \text{ and } x \in \mathsf{X}.$$

Remark 6. Notice from the proof that the moment condition is assumed here in order to simplify our statement and that more general conditions are possible.

Proof. For any $x \in \mathsf{X}$ let $A_\mathsf{Z}(x) := \{z \in \mathsf{Z} : \pi(x+z)/\pi(x) \geq 1\}$ and $R_\mathsf{Z}(x) := A^c_\mathsf{Z}(x)$ (where the complement is with respect to Z) and $A(x) := x + A_\mathsf{Z}(x)$. Without loss of generality we focus on the case $x > 0$. From Lemma 2 there exists $R_1 > 0$ such that for any $x \geq R_1$, $R_\mathsf{Z}(x) = (-\infty, -x + \Upsilon(x)) \cup (0, \infty)$ and $A_\mathsf{Z}(x) = [-x + \Upsilon(x), 0]$, where $\Upsilon(x)$ is as in Lemma 2. For, $x \geq R_1$ and $\sigma \leq 1$,we have the inequalities

$$\begin{aligned}
\alpha_\sigma(x) &= \int_\mathsf{Z} \min\left\{1, \frac{\pi(x+z)}{\pi(x)}\right\} q_\sigma(z)dz \\
&= 1 + \int_{R_\mathsf{Z}(x)} \left[\frac{\pi(x+z)}{\pi(x)} - 1\right] q_\sigma(z)dz \\
&\geq 1 - \int_{R_\mathsf{Z}(x)} q_\sigma(z)dz \\
&= \frac{1}{2} - \int_{-\infty}^{(-x+\Upsilon(x))/\sigma} q(z)dz \\
&\geq \frac{1}{2} - \int_{-\infty}^{-x/\sigma} q(z)dz \quad .
\end{aligned}$$

Now with $\mu_1 < \infty$ the first order moment of q we notice that from Chebyshev's inequality and for $x \geq R_1$

$$\int_{x/\sigma}^\infty q(z)dz \leq \sigma R_1^{-1} \times \mu_1/2$$

from which we deduce the first statement for $\sigma \leq 1$ and $x \geq R_1$. Now for $x \geq R_1$ and $\sigma \geq 1$

$$\begin{aligned}
\alpha_\sigma(x) &\leq \frac{\bar{q}}{\sigma}\left(\int_{A_\mathsf{Z}(x)\cup R_\mathsf{Z}(x)} \min\left\{1, \frac{\pi(x+z)}{\pi(x)}\right\} dz\right) \\
&\leq \frac{\bar{q}}{\sigma}\left(2C_{\Upsilon,1}\left(-\log(\pi(x)/C_{\Upsilon,2})\right)^{1/p} + \int_0^\infty \frac{\pi(\Upsilon(x)-z)}{\pi(\Upsilon(x))}dz + \int_0^\infty \frac{\pi(x+z)}{\pi(x)}dz\right). \\
&\leq C\frac{(-\log(\pi(x)/C_{\Upsilon,2}))^{1/p}}{\sigma} \quad ,
\end{aligned}$$

where we have used the results of Lemma 2 to upper bound the Lebesgue measure of $A_\mathsf{Z}(x)$ and the last two integrals. We now turn to the case $0 \leq x \leq R_1$. Let $M > 0$ such that $\int_M^\infty q(z)dz \leq 1/4$ and $\sigma \leq 1$ and with

$\phi_x(z) = \pi(x+z)/\pi(x)$

$$
\begin{aligned}
\alpha_\sigma(x) &= 1 + \frac{1}{\sigma}\int_{\mathsf{Z}} \left(1 \wedge \frac{\pi(x+z)}{\pi(x)} - 1\right) q\left(\frac{z}{\sigma}\right) dz \\
&\geq 1 + \int_{-\infty}^{-M} \left(1 \wedge \frac{\pi(x+\sigma z)}{\pi(x)} - 1\right) q(z)dz + \int_{M}^{\infty} \left(1 \wedge \frac{\pi(x+\sigma z)}{\pi(x)} - 1\right) q(z)dz - \int_{-M}^{M} \left|\frac{\pi(x+\sigma z)}{\pi(x)} - 1\right| q(z)dz \\
&\geq 1 - 2\int_M^\infty q(z)dz - 2M\bar{q} \sup_{x\in B(0,R_1), z \in B(0,M)} |\phi'_x(z)|\,\sigma
\end{aligned}
$$

and we deduce the first statement of the lemma. We now consider the case $0 \leq x \leq R_1$ and $\sigma \geq 1$. There exists (cf. the proof of Lemma 2) $R_2 > 0$ such that for all $x \leq R_1$

$$
\alpha_\sigma(x) \leq \sigma^{-1}\int_{-R_2}^{R_2} q(z/\sigma)dz + \sigma^{-1}\int_{-\infty}^{-R_2} \frac{\pi(x+z)}{\pi(x)} q(z/\sigma)dz + \sigma^{-1}\int_{R_2}^{\infty} \frac{\pi(x+z)}{\pi(x)} q(z/\sigma)dz \quad .
$$

From the proof of Lemma 2, we have the bound $\pi(x+z)/\pi(x) \leq C_1 \exp\left(-C_\ell/p|x - |z||^p\right)$ and since $q(z) \leq \bar{q}$ we deduce the existence $C > 0$ such that for $x \leq R_1$ and $\sigma \geq 1$ we have $\alpha_\sigma(x) \leq C/\sigma$. □

Remark 7. Note that the restriction $\alpha_* \in (0, 1/2)$ is practically harmless since this covers relevant values according to the scaling theory of the RWM [8].

5.3.2 Proof for the accelerated version: (ALG4).

The arguments are similar to those of Subsection 5.3.1 but $w(\theta)$ is here of a different form.

(Proof of Theorem 6 in the case of (ALG4)). The beginning of the proof is similar to that for (ALG3) up to the definition of $w(\theta)$. Here we choose $w(\theta) = 1 + |\theta|^2$ and from Lemma 6 we have, since for $x \geq 0$ $\exp(x) \geq 1 + x^2$, the existence of $C > 0$ such that

$$
\begin{aligned}
\alpha_{\exp(\theta)}(x) &\geq 1/2 - C\frac{V^\beta(x)}{1+|\theta|^2} && \text{for } \theta \leq 0 \ \text{ and } x \in \mathsf{X} \\
\alpha_{\exp(\theta)}(x) &\leq C\frac{V^\beta(x)}{1+|\theta|^2} && \text{for } \theta \geq 0 \ \text{ and } x \in \mathsf{X} \quad .
\end{aligned}
$$

The proof is now identical to that for (ALG3) by using Proposition 6. □

Proposition 6. *Consider (ALG4) with $\alpha_* \in (0, 1/2)$, let $w(\theta) := 1+|\theta|^2$ and assume that there exists $C > 0$ and $\beta \in [0, 1)$ such that for any $\theta, x \in \Theta \times \mathsf{X}$*

$$
\operatorname{sgn}(\theta)(\alpha_{\exp(\theta)}(x) - \alpha_*) \leq -[\alpha_* \wedge (1/2 - \alpha_*)] + CV^\beta(x)/w(\theta) \quad .
$$

Let $\gamma_{\max} \in (0, \alpha_* \wedge (1/2 - \alpha_*))$. Then there exists $C' > 0$ such that for any $\gamma \in (0, \gamma_{\max}]$ and $\theta, x \in \Theta \times \mathsf{X}$,

$$
P_{\theta,\gamma} w(\theta, x) \leq w(\theta) - \gamma w(\theta)\Delta\left(\mathbb{I}\{|\theta| \leq 1\}C'^{-1}\left(\alpha_* \wedge (1/2 - \alpha_*)\right) + V^\beta(x)/w(\theta)\right)
$$

with

$$
\Delta(z) = 2[\alpha_* \wedge (1/2 - \alpha_*) - \gamma_{\max} - C'z] \quad .
$$

Proof. With $\theta_+ = \theta + \gamma(1 + |\theta|)[\alpha(x, y) - \alpha_*]$ we have

$$
\begin{aligned}
w(\theta_+) &\leq w(\theta) + 2\gamma|\theta|(|\theta| + 1)\operatorname{sgn}(\theta)(\alpha(x, y) - \alpha_*) + \gamma^2(1 + |\theta|)^2[\alpha(x, y) - \alpha_*]^2 \\
&\leq w(\theta) + 2\gamma|\theta|(|\theta| + 1)\operatorname{sgn}(\theta)(\alpha(x, y) - \alpha_*) + 2\gamma^2 w(\theta)
\end{aligned}
$$

so

$$
P_{\theta,\gamma} w(\theta, x) \leq w(\theta) + 2\gamma|\theta|(|\theta| + 1)\left[-[\alpha_* \wedge (1/2 - \alpha_*)] + CV^\beta(x)/w(\theta)\right] + 2\gamma^2 w(\theta)
$$

Notice that for $|\theta| \geq 1$ we have $|\theta|(1 + |\theta|) \geq 1 + |\theta|^2$. Consequently for any $|\theta| \geq 1$ and $x \in \mathsf{X}$ such that $-[\alpha_* \wedge (1/2 - \alpha_*)] + CV^\beta(x)/w(\theta) \leq 0$

$$
\begin{aligned}
P_{\theta,\gamma} w(\theta, x) &\leq w(\theta) + 2\gamma w(\theta)\left[-[\alpha_* \wedge (1/2 - \alpha_*)] + \gamma + CV^\beta(x)/w(\theta)\right] \\
&\leq w(\theta) - 2\gamma w(\theta)\left[\alpha_* \wedge (1/2 - \alpha_*) - \gamma - 2CV^\beta(x)/w(\theta)\right] \quad .
\end{aligned}
$$

Notice that for any $\theta \in \Theta$, $|\theta|(1+|\theta|) \leq (1+|\theta|)^2 \leq 2(1+|\theta|^2)$. For the specific case $-[\alpha_* \wedge (1/2-\alpha_*)] + CV^\beta(x)/w(\theta) \geq 0$ we therefore have

$$\begin{aligned}P_{\theta,\gamma} w(\theta,x) &\leq w(\theta) + 2\gamma 2 w(\theta)\left[-[\alpha_* \wedge (1/2-\alpha_*)] + \gamma/2 + CV^\beta(x)/w(\theta)\right] \\ &\leq w(\theta) + 2\gamma w(\theta)\left[-[\alpha_* \wedge (1/2-\alpha_*)] + \gamma + 2CV^\beta(x)/w(\theta)\right]\end{aligned}$$

and for any $\theta, x \in \Theta \times \mathsf{X}$ one has

$$P_{\theta,\gamma} w(\theta,x) \leq w(\theta) - 2\gamma w(\theta)[-2CV^\beta(x)/w(\theta) - \gamma] \quad .$$

We can now combine these intermediate results, yielding for any $\theta, x \in \Theta \times \mathsf{X}$

$$P_{\theta,\gamma} w(\theta,x) \leq w(\theta) - \gamma 2 w(\theta)\left[\alpha_* \wedge (1/2-\alpha_*) - \gamma_{\max} - \mathbb{I}\{|\theta| \leq 1\}\left(\alpha_* \wedge (1/2-\alpha_*)\right) - 2CV^\beta(x)/w(\theta))\right] \quad ,$$

and we conclude. □

6 Acknowledgements

The first author was partially funded by an EPSRC Advance Research Fellowship and a Winton Capital research prize for this research. The third author was supported by the Academy of Finland (project 250575) and by the Finnish Academy of Science and Letters, Vilho, Yrjö and Kalle Väisälä Foundation.

A Appendix

Lemma 7. *Let $c, p > 0$ be constants. Then, there exists constants $M = M(c,p) \in (0,\infty)$ and $x_0 = x_0(c,p) \in (0,\infty)$ such that*

$$\int_0^\infty \exp\big(-c[(x+z)^p - x^p]\big)dz \leq Mx^{1-p} \qquad \textit{for all } x \geq x_0 \ .$$

Proof. By a change of variable $u = c(x+z)^p$, we obtain

$$\int_0^\infty \exp\big(-c[(x+z)^p - x^p]\big)dz = \frac{e^{cx^p}}{cp}\int_{cx^p}^\infty e^{-u}u^{\frac{1}{p}-1}du \ .$$

Integration by parts yields

$$\int_{cx^p}^\infty e^{-u}u^{\frac{1}{p}-1}du = e^{-cx^p}(cx^p)^{\frac{1}{p}-1} + \left(\frac{1}{p}-1\right)\int_{cx^p}^\infty e^{-u}u^{\frac{1}{p}-2}du \ . \tag{A.1}$$

Now, if $p \geq 1$, this is enough to yield the claim. Suppose then $p \in (0,1)$, and fix a constant $\lambda \in (0,1)$. By (A.1),

$$(1-\lambda)\int_{cx^p}^\infty e^{-u}u^{\frac{1}{p}-1}du = e^{-cx^p}(cx^p)^{\frac{1}{p}-1} + \int_{cx^p}^\infty e^{-u}u^{\frac{1}{p}-1}\left[\left(\frac{1}{p}-1\right)\frac{1}{u} - \lambda\right]du \ .$$

Now, if $cx^p \geq \left(\frac{1}{p}-1\right)\frac{1}{\lambda}$, the latter integrand is negative. Setting

$$x_0 := \left[\left(\frac{1}{p}-1\right)\frac{1}{\lambda c}\right]^{1/p} ,$$

we therefore have for $x \geq x_0$ the desired bound

$$\int_0^\infty \exp\big(-c[(x+z)^p - x^p]\big)dz \leq \frac{(cx^p)^{\frac{1}{p}-1}}{cp(1-\lambda)} = \frac{c^{\frac{1}{p}-2}}{p(1-\lambda)}x^{1-p} \ .$$

We remark that the constant $\lambda \in (0,1)$ can be used to optimise the values constants M and x_0. □

B Proofs from Section 3

We state the following result for the reader's convenience.

Theorem 7 (see [3] for a proof)**.** *Assume (A5). For any $M \in (M_0, M_1]$ there exist $\delta_0 > 0$ and $\lambda_0 > 0$ such that, for all $n \geq 1$, all $\vartheta_0 \in \mathcal{W}_{M_0}$, all sequences $\rho = \{\rho_k\}$ of non negative real numbers and all sequences $\{\varsigma_k\} \subset \Theta^{\mathbb{N}}$ of n_θ-dimensional vectors satisfying*

$$\sup_{1 \leq k \leq n} \rho_k \leq \lambda_0 \quad \textit{and} \quad \sup_{1 \leq k \leq n} \left| \sum\nolimits_{j=1}^{k} \rho_j \varsigma_j \right| \leq \delta_0,$$

we have for $k = 1, \ldots, n$, $w(\vartheta_k) \leq M$, where $\vartheta_k = \vartheta_{k-1} + \rho_k h(\vartheta_{k-1}) + \rho_k \varsigma_k$.

References

[1] Y. F. Atchadé and G. Fort, Limit Theorems for some adaptive MCMC algorithms with subgeometric kernels, Bernoulli Volume 16, Number 1 (2010), 116-154, 2010.

[2] Andrieu C. and Moulines É., "On the ergodicity properties of some adaptive MCMC algorithms," *Annals of Applied Probability*, Vol. 16, No. 3, pp. 1462-1505, 2006.

[3] Andrieu C., Moulines É. and Priouret P., "Stability of stochastic approximation under verifiable conditions", *SIAM Journal on Control and Optimisation*, Vol. 44, Number 1, pp. 283-312, 2005.

[4] Andrieu C. and Robert, C.P., "Controlled MCMC for optimal sampling," *Tech. Rep.* 0125, Cahiers de Mathématiques du Ceremade, Université Paris-Dauphine, 2001.

[5] Andrieu C. and Vihola, M., "Markovian stochastic approximation with expanding projections," Univ. of Bristol, Tech. Report, 2011.

[6] Y. F. Atchadé and J. S. Rosenthal, "On Adaptive Markov Chain Monte Carlo Algorithms," *Bernoulli* 11, 815-828, 2005.

[7] Delyon, B. and Juditsky, A. Accelerated stochastic approximation, *SIAM Journal on Optimization*, vol. 3, pp. 868-881, 1993.

[8] Gelman, A., Roberts, G., and Gilks, W., "Efficient Metropolis Jumping Rules", In Bayesian Statistics 5. Oxford University Press, New York, 1995.

[9] Haario, H., Saksman, E. and Tamminen, J., "Adaptive Proposal Distribution for Random Walk Metropolis Algorithm", *Computational Statistics*, vol. 14, no. 3, pp 375-395, 1999.

[10] Haario, H., Saksman, E. and Tamminen, J., "An Adaptive Metropolis Algorithm", *Bernoulli*, vol. 7, no. 2, 2001.

[11] Jarner, S. F. and Hansen, E. (1998), "Geometric Ergodicity of Metropolis algorithms", *Stoch. Proc. Appl.*, **85**, 341-161.

[12] Meyn, S.P. and Tweedie, R.L. (1993), *Markov Chains and Stochastic Stability*, Springer-Verlag.

[13] Roberts, G.O. and Rosenthal, J.S., "Coupling and Ergodicity of Adaptive MCMC," *J. Appl. Probab.,* vol 44, no. 2 , pp. 458-475, 2007.

[14] E. Saksman and M. Vihola. On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.*, 20(6):2178–2203, Dec. 2010.

[15] L. Younes, On the convergence of Markovian stochastic algorithms with rapidly decreasing ergodicity rates, Stochastics and Stochastics Reports, 65, 177-228 (1999).

[16] A. Benveniste, M. Métivier, and P. Priouret : Adaptive algorithms and stochastic approximations. Springer Verlag, Applications of Mathematics, vol. 22, Berlin, Heidelberg, New York, 1990 (also available in French, Masson 1987).

[17] M. Vihola. On the stability and ergodicity of adaptive scaling Metropolis algorithms. Stochastic Process. Appl., 121(12):2839–2860, 2011.